Fast and Smooth Interpolation on Wasserstein Space

Sinho Chewi   **Julien Clancy**   **Thibaut Le Gouic**

Philippe Rigollet   **George Stepaniants**   **Austin J. Stromme**

Massachusetts Institute of Technology
{schewi, julienc, tlegouic, rigollet, gstepan, astromme}@mit.edu

Abstract

We propose a new method for smoothly interpolating probability measures using the geometry of optimal transport. To that end, we reduce this problem to the classical Euclidean setting, allowing us to directly leverage the extensive toolbox of spline interpolation. Unlike previous approaches to measure-valued splines, our interpolated curves (i) have a clear interpretation as governing particle flows, which is natural for applications, and (ii) come with the first approximation guarantees on Wasserstein space. Finally, we demonstrate the broad applicability of our interpolation methodology by fitting surfaces of measures using thin-plate splines.

1 INTRODUCTION

Smooth interpolation is a fundamental tool in numerical analysis that plays a central role in data science. While this task is traditionally studied on the flat Euclidean space $\mathbb{R}^d$, recent applications have called for interpolation of points living on curved spaces such as smooth manifolds (Noakes, Heinzinger, and Paden, 1989) and, more recently, the Wasserstein space of probability measures. An important application arises in single-cell genomic data analysis where the measure $\mu_t^\star$ represents a population of cells at time t of a biological process such as differentiation, and the cells of an organism specialize over the course of early development. In this context, two main questions arise: 1) to infer the profile of the population at unobserved
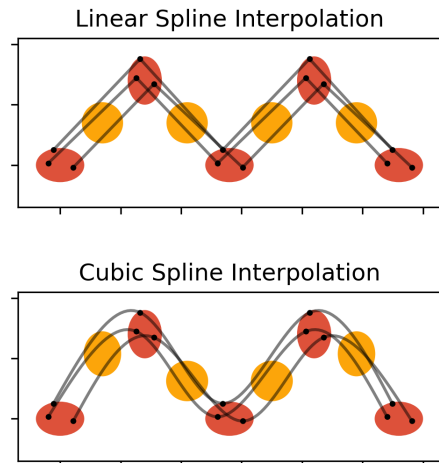

Figure 1: Piecewise linear and cubic spline interpolation of four Gaussians. The interpolation knots are shown in red and the interpolated Gaussians are shown in orange. See Appendix E.1.

times; and more importantly 2) to reconstruct the trajectories of individual cells in gene space, that is: given a cell at time t, determine its (likely) history and fate. Regev et al. (2017) argue that cellular trajectory reconstruction is crucial to unlocking the promises of single-cell genomics. A breakthrough in this direction was recently achieved using optimal transport by Schiebinger et al. (2019), but their work does not produce *smooth* trajectories. To illustrate, we display in Figure 1 a comparison of their approach with the smooth interpolation methodology developed in the present work. Although we are mainly motivated by cell trajectory reconstruction, we are confident that the flexibility and efficiency of the method will allow it to find applications beyond this scope.

While the first question above is a natural extension

of interpolation to the space of probability measures, the second question calls for a specific type of interpolation: one that also reconstructs the (smooth) trajectories of individual particles. Mathematically, the trajectory of a particle (e.g., a cell) is a stochastic process $(X_t^\star)_{t \in [0,1]}$ with smooth sample paths. This leads us to the following problem of trajectory-aware interpolation over the space of probability measures.

The problem. Let $(X_t^\star)_{t \in [0,1]}$ be a stochastic process on $\mathbb{R}^d$ with $\mathcal{C}^2$ sample paths and marginal laws $X_t^\star \sim \mu_t^\star, t \in [0,1]$. Given $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$ at times $0 = t_0 < t_1 < \cdots < t_N = 1$, construct a stochastic process $(X_t)_{t \in [0,1]}$ such that X_t has $\mathcal{C}^2$ sample paths and the distribution μ_t of X_t *interpolates* the given measures, meaning $\mu_{t_i} = \mu_{t_i}^\star$ for $i = 0, 1, \ldots, N$.

Throughout, we assume all given measures to be absolutely continuous with finite second moment, and (as advocated in Schiebinger et al. (2019)) we equip this space with the 2-Wasserstein metric W_2 and seek an interpolation that reflects this geometry.

Prior work. This work is at the intersection of interpolation and optimal transport. On the one hand, interpolation in $\mathbb{R}^d$ is very well-developed, with fast and accurate methods ranging from interpolating polynomials and splines to more exotic non-parametric approaches (Wahba, 1990), and with renewed interest due to recent theoretical results (Belkin et al., 2019). Our methodology can accommodate all of these options, but we focus on cubic spline interpolations due to their simplicity, theoretical guarantees, and their *curvature-minimizing* property (see Section 3). On the other hand, optimal transport has become a useful tool in the analysis of observations represented in the form of probability measures. Recent computational advances (Cuturi, 2013; Altschuler, Weed, and Rigollet, 2017; Peyré and Cuturi, 2019) have led to the development of many methods in statistical optimal transport, from barycenters to geodesic PCA. The present work extends this toolbox by developing a method for smooth interpolation over the Wasserstein space of probability measures.

Splines in Wasserstein space were considered concurrently and independently by Chen, Conforti, and Georgiou (2018) and Benamou, Gallouët, and Vialard (2019). Both papers converge to the same notion of splines, which we call P-splines. Though motivated by particle dynamics, P-splines solve an optimal transport problem that is not guaranteed to have a *Monge* solution. Instead, it outputs stochastic processes $(X_t)_{t \in [0,1]}$ for which X_t is not a deterministic function of X_0. In other words, given an initial position, there is no unique particle trajectory emanating from this position but rather a superposition

of such trajectories; see Figure 2 and the discussion in Section 3. We show that this is not an isolated phenomenon arising from pathological data but applies even to the canonical example of one-dimensional Gaussian distributions. This limitation, together with a relatively heavy computational cost, severely hinders the deployment of P-splines in applications, ours included, especially where interpretation is a priority.

We review these prior works and their motivations in Section 3. We remark however that the algorithm we ultimately propose requires considerably less technical machinery to describe compared to these prior works, and we recommend that readers who simply wish to understand our method skip directly to Section 4.1.

Our contributions. To overcome the aforementioned issues, we propose in Section 4.1 a new method for constructing measure-valued splines. Our method outputs Monge solutions, and moreover enjoys significant computational advantages: it only requires N evaluations of Monge maps and standard Euclidean cubic spline fitting to output trajectories. In the case where all of the measures are Gaussian, our approach is more interpretable and scalable than the SDP-based approach of Chen, Conforti, and Georgiou (2018).

In particular, for Gaussian measures, our method only requires one $d \times d$ matrix inversion and $O(1)$ multiplications per sample point $\mu_{t_i}^\star$. In comparison, the method of Chen, Conforti, and Georgiou (2018) solves an SDP with N coupled $4d \times 4d$ matrix variables. In the general case we still only need to perform N pairwise OT computations, which can be done efficiently (Altschuler, Weed, and Rigollet, 2017), while the competing algorithms in Benamou, Gallouët, and Vialard (2019) require time exponential in either N or d.

Our new method comes with a theoretical study of its approximation error. In the Gaussian setting, we introduce new techniques for studying quantitative approximation of transport maps and vector fields. In turn, it yields an approximation guarantee analogous to the classical setting (Theorem 2), but adapted to the geometry of the space. This paves the way for a principled theory of approximation on Wasserstein space that mirrors classical Euclidean results. In a forthcoming work, we build upon these ideas to develop higher-order approximation schemes.

A key feature of our approach is its flexibility, which allows us to easily extend our method to fitting thin-plate splines for measures indexed by high-dimensional covariates. We study the case of two-dimensional spatial covariates in Section 6.

Notation. For a curve such as $(\mu_t)_{t \in [0,1]}$ or $(X_t)_{t \in [0,1]}$, defined over $[0,1]$, we use the concise notations (μ_t)
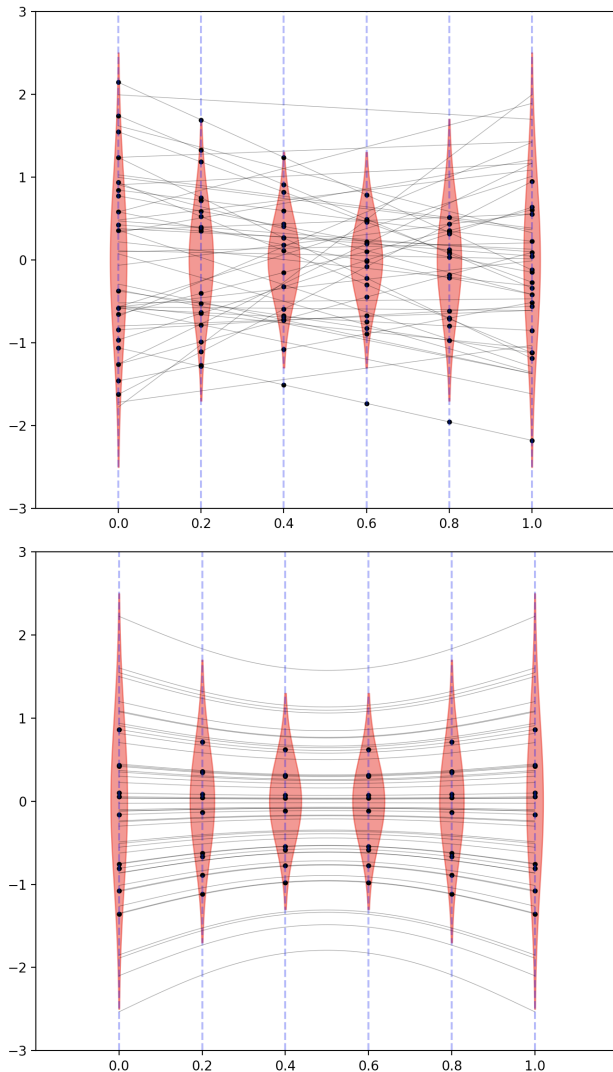
and (X_t) respectively, where the time variable t is always understood to range over the interval $[0, 1]$.

2 BACKGROUND ON OPTIMAL TRANSPORT

In this section, we recall useful notions from optimal transport and provide some of the key theory used for Wasserstein splines. We refer readers to the standard textbooks Villani (2003), Villani (2009), and Santambrogio (2015) for introductory treatments.

Given two probability measures μ_0, μ_1 on $\mathbb{R}^d$ with finite second moment, the 2-Wasserstein distance W_2 is defined as

$$W_2^2(\mu_0, \mu_1) := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \|x - y\|^2 \, \mathrm{d}\pi(x, y), \quad (1)$$

where $\Pi(\mu_0, \mu_1)$ is the set of all joint distributions with marginals μ_0 and μ_1. This indeed defines a distance on probability measures with finite second moment, and we denote the resulting metric space by $\mathcal{P}_2(\mathbb{R}^d)$. If μ_0 has a density with respect to Lebesgue measure, then the solution of (1) is unique, and it is supported on the graph of a function $T : \mathbb{R}^d \to \mathbb{R}^d$, called the *Monge map*. Moreover, it is characterized as the unique mapping such that (i) the pushforward of μ_0 via T is μ_1 and (ii) there exists a convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \infty$ such that $T = \nabla\phi$. That is, if $X_0 \sim \mu_0$, the solution of (1) is the law of $(X_0, \nabla\phi(X_0))$. For the rest of the paper, without further comment, we work exclusively with probability measures that admit a density and have a finite second moment.

It has been understood since the seminal work of Otto that $\mathcal{P}_2(\mathbb{R}^d)$ exhibits many of the properties of a Riemannian manifold, a fact which has been instrumental to applications of optimal transport to partial differential equations (Jordan, Kinderlehrer, and Otto, 1998; Carrillo and Vaes, 2019), sampling (Bernton, 2018; Durmus, Majewski, and Miasojedow, 2019; Y. Lu, J. Lu, and Nolen, 2019; Chewi et al., 2020a; Chewi et al., 2020c), and barycenters (Backhoff-Veraguas et al., 2018; Zemel and Panaretos, 2019; Chewi et al., 2020b). Specifically, given a regular curve (μ_t), there is a well-defined notion of a "tangent vector" v_t to the curve at time t. This is a vector *field* of instantaneous particle velocities, where μ_t is interpreted as the law of the particles at time t. The field v_t arises from optimally coupling the curve at nearby times, and we have the limiting result

$$v_t = \lim_{h \to 0} \frac{T_{\mu_t \to \mu_{t+h}} - \mathrm{id}}{h} \qquad \text{in } L^2(\mu_t) \qquad (2)$$

where $T_{\mu_t \to \mu_{t+h}}$ is the Monge map between μ_t and μ_{t+h}. For a proof see Ambrosio, Gigli, and Savaré (2008, Proposition 8.4.6).

Figure 2: A comparison of 50 trajectories sampled from P-splines and transport splines for the Gaussian interpolation problem in Proposition 1 (see Appendix A.1 for a detailed discussion). The first figure shows trajectories drawn from the P-spline interpolation, while the second shows trajectories from our method.

This differential structure has been especially useful in fluid dynamics, by connecting the equivalent Eulerian and Lagrangian perspectives on particle flows. The former keeps track of the density μ_t and velocity v_t of particles passing through any given time and spatial position. In contrast, the Lagrangian perspective tracks the trajectories of individual particles, which can be obtained as integral curves of the velocity fields; that is, we solve the ODE

$$\dot{X}_t = v_t(X_t), \qquad X_0 \sim \mu_0.$$

Chosing the vector fields v_t to be the tangent vectors above precisely yields that $X_t \sim \mu_t$. Thus, the Lagrangian perspective associates a natural stochastic process, (X_t), with the curve of measures (μ_t); we therefore refer to the process (X_t) as the *Lagrangian coupling*. See Villani (2003, §5.4) for further details.

3 SPLINES ON EUCLIDEAN SPACE, MANIFOLDS, AND WASSERSTEIN SPACE

We recall the definition of *natural cubic splines*. Given points $(x_0, x_1, \ldots, x_N) \subset \mathbb{R}^d$ to interpolate at a sequence of times $0 = t_0 < t_1 < \cdots < t_N = 1$, consider the variational problem

$$\min_{(\gamma_t)} \int_0^1 \|\ddot{\gamma}_t\|^2 \, \mathrm{d}t \quad \text{s.t.} \quad \gamma_{t_i} = x_i \text{ for all } i. \quad (3)$$

The solution to this minimization problem is a piecewise cubic polynomial that is globally $\mathcal{C}^2$ and has zero acceleration at times $t_0 = 0$ and $t_N = 1$.

Based on this energy-minimizing property, there is a natural generalization of cubic splines to Riemannian manifolds: in (3) the acceleration $\ddot{\gamma}$ is replaced with its Riemannian analogue, the covariant derivative $\nabla_{\dot{\gamma}} \dot{\gamma}$ of the velocity, and the norm $\|\cdot\|$ is given by the Riemannian metric. However, unlike its Euclidean counterpart, there is no general algorithm to fit Riemannian cubic splines, leading to alternative proposals (Gousenbourger, Massart, and Absil, 2019).

In addition to a first-order differentiable structure (the tangent space), Gigli (2012) has developed a second-order calculus on $\mathcal{P}_2(\mathbb{R}^d)$, including a covariant derivative ∇. Thus, in analogy with the Riemannian setting, we can define *energy splines* (*E-splines* in short) via

$$\inf_{(\mu_t, v_t)} \int_0^1 \|\nabla_{v_t} v_t\|_{L^2(\mu_t)}^2 \, \mathrm{d}t \quad \text{s.t.} \quad \mu_{t_i} = \mu_{t_i}^\star \text{ for all } i \quad (4)$$

where the minimization is taken over all curves (μ_t) and their tangent vectors (v_t) (see Section 2). The solution to this problem naturally yields a stochastic process (X_t) with marginal laws (μ_t), namely: we draw $X_0 \sim \mu_0$, and conditioned on X_0 the rest of the trajectory is determined by the ODE $\dot{X}_t = v_t(X_t)$.

E-splines were introduced concurrently by Chen, Conforti, and Georgiou (2018) and Benamou, Gallouët, and Vialard (2019). Since E-splines are intractable, these authors proposed a relaxation which we call *path splines* (*P-splines* in short):

$$\inf_{(X_t)} \int_0^1 \mathbb{E}[\|\ddot{X}_t\|^2] \, \mathrm{d}t, \quad (5)$$

where the infimum is taken over stochastic processes (X_t) with values in $\mathbb{R}^d$ and such that $X_{t_i} \sim \mu_{t_i}^\star$ for all $i = 0, 1, \ldots, N$. (This is indeed a relaxation in a formal sense detailed in the papers referenced above.) The name derives from the fact that this is an optimization over measures in path space, and the problem (5) can be reduced to a multimarginal optimal transport problem with quadratic cost.

Unfortunately, though solvable in principle, the formulation (5) remains difficult to compute and its solution is not necessarily induced by a deterministic map; that is, there is no guarantee of a deterministic function $\phi_t : \mathbb{R}^d \to \mathbb{R}^d$ such that $X_t = \phi_t(X_0)$. This point is particularly problematic for inference of trajectories as illustrated in Figure 2.

Given the various definitions of splines, some natural questions arise. Specifically, the papers above left open the question of whether E-splines coincide with P-splines, and whether the solution to the P-spline problem is necessarily induced by Monge maps. We conclude this section by resolving these questions in the negative.

Proposition 1 (informal). *There exist Gaussian data $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$ such that there is a unique jointly Gaussian solution to the P-spline problem (5) and it is* not *induced by a deterministic map.*

Proposition 2 (informal). *There exist Gaussian data $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$ for which the E-spline (4) and P-spline (5) interpolations do not coincide.*

Investigation of these questions requires some care, since there are many subtleties regarding the definitions. We give a careful discussion and proofs in Appendix A.

4 TRANSPORT SPLINES

4.1 The Algorithm

To address the difficulties discussed in the previous section, we propose a new method for measure interpolation, which we call *transport splines*. Our framework decouples the interpolation problem into two steps:

1. Couple the given measures, that is, construct a random vector $(X_{t_0}, X_{t_1}, \ldots, X_{t_N})$ with marginal laws $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$.

2. Apply a Euclidean interpolation algorithm to the points $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$.

A convenient choice for the second step is to use cubic splines, but our framework works equally well with other standard Euclidean methods and can be adapted to the application at hand. We illustrate this point in Section 6, where we construct surfaces interpolating one-dimensional measures using thin-plate splines.

A simple and practical choice for the first step, which we explore in the present paper, is to couple the random variables $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$ successively using the Monge maps between them. That is, we draw $X_{t_0} \sim \mu_{t_0}^\star$, and for each $i = 1, \ldots, N$ we set $X_{t_i} = T_i(X_{t_{i-1}})$, where T_i is the Monge map from $\mu_{t_{i-1}}^\star$ to $\mu_{t_i}^\star$. The second step then reduces to interpolating $X_{t_0}, T_1(X_{t_0}), \ldots, T_N \circ \cdots \circ T_1(X_{t_0})$ in Euclidean space. The interpolation property of transport splines follows readily from the definition of Monge maps since $T_i \circ \cdots \circ T_1(X_{t_0}) \sim \mu_{t_i}^\star$.

For the task of outputting sample trajectories from the transport spline, we summarize our method in Algorithm 1, and we display an application to the reconstruction of trajectories in a many-body physical system in Figure 3. In the next section, we provide detailed motivation for the first step of the algorithm which builds on background from Sections 2 and 3.

Algorithm 1 Sample Transport Spline Trajectories

1: **procedure** INTERPOLATE($(t_i)_{i=0}^N$, $(\mu_{t_i}^\star)_{i=0}^N$)
2:     Draw $X_{t_0} \sim \mu_{t_0}^\star$
3:     **for** $i = 1, \ldots, N$ **do**
4:         Set $X_{t_i} = T_i(X_{t_{i-1}})$, where T_i is the Monge map from $\mu_{t_{i-1}}^\star$ to $\mu_{t_i}^\star$
5:     **end for**
6:     Interpolate the points $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$ to obtain a curve (X_t)
7:     **output** (X_t)
8: **end procedure**

4.2 Motivation

The choice of coupling in the first step of our method is motivated by the geometry of $\mathcal{P}_2(\mathbb{R}^d)$. If the observations $\mu_{t_0}^\star, \ldots, \mu_{t_N}^\star$ sit along a curve of measures $(\mu_t^\star)$, then (as discussed in Section 2) there is an associated Lagrangian coupling $(X_t^\star)$ satisfying $\dot{X}_t^\star = v_t^\star(X_t^\star)$. Thus if $\delta = t_1 - t_0$, then $X_{t_1}^\star = X_{t_0}^\star + \delta v_{t_0}^\star(X_{t_0}^\star) + o(\delta)$. On the other hand, from (2) the Monge map T_1 gives a

N-Body Natural Cubic Spline Interpolation

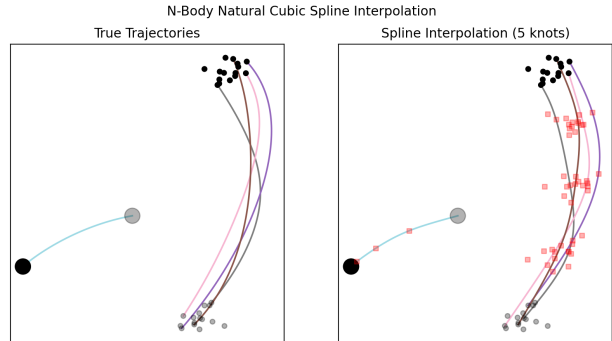True Trajectories   Spline Interpolation (5 knots)

Figure 3: Reconstruction of trajectories in a physical system. See Appendix E.2.

first-order approximation to $v_{t_0}^\star$: $T_1 - \mathrm{id} = \delta v_{t_0}^\star + o(\delta)$ (see Ambrosio, Gigli, and Savaré (2008, Proposition 8.4.6)). Combining these approximations we get $T_1(X_0^\star) = X_{t_1}^\star + o(\delta)$. From this heuristic discussion, one expects that as the mesh size $\max_{i \in [N]}(t_i - t_{i-1})$ tends to zero, the coupling $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$ obtained via successive Monge maps is a good approximation to the Lagrangian coupling $(X_{t_0}^\star, X_{t_1}^\star, \ldots, X_{t_N}^\star)$.

4.3 Relationship with E-Splines in One Dimension

Although E-splines are in general intractable, in the one-dimensional case it turns out that there are many situations of interest in which E-splines coincide with transport splines. Indeed, suppose that the measures $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$ are all one-dimensional, and for a measure μ let $F_\mu^\dagger$ denote its quantile function.[1] Let (G_t) be the natural cubic spline in $L^2[0,1]$ interpolating the quantile functions $F_{\mu_{t_0}^\star}^\dagger, F_{\mu_{t_1}^\star}^\dagger, \ldots, F_{\mu_{t_N}^\star}^\dagger$. Then:

Theorem 1. *Suppose that for all t, G_t is a valid[2] quantile function. Then the transport spline and the E-spline* (4) *both coincide with the curve* (μ_t) *where μ_t has quantile function G_t. Furthermore, if (X_t) is the stochastic process associated with the transport spline and $(X_t^\star)$ is the Lagrangian coupling for the E-spline, then (X_t) and $(X_t^\star)$ have the same distribution as the law of $(G_t(U))$, where U is a uniform random variable on* $[0, 1]$.

We emphasize that, in light of the counterexamples described at the end of Section 3, the P-spline and E-

---

[1]Under our assumption that the measures are absolutely continuous, the quantile function $F_\mu^\dagger$ simply coincides with the inverse CDF F_μ^{-1}, but we use the quantile function notation here to reflect the general embedding $\mathcal{P}_2(\mathbb{R}) \hookrightarrow L^2[0,1]$.

[2]A valid quantile function $G_t : [0,1] \to \mathbb{R} \cup \{\pm\infty\}$ is increasing and right-continuous.

spline are likely to differ generically and, in fact, they differ in the Gaussian case, which is covered by the above theorem (see Appendix A.2). Therefore, it appears that the transport spline is more suitable as a relaxation of the E-spline when interpolating univariate distributions. We also emphasize that these results do not affect the applicability of our method in dimensions higher than one.

We give the proof of Theorem 1 in Appendix B.

5 THE GAUSSIAN CASE

We now focus on the Gaussian case and we assume that we employ natural cubic splines in Step 2 of our algorithm. For simplicity, we can assume that the measures are centered.[3] A centered non-degenerate Gaussian can be identified with its covariance matrix, and the Wasserstein distance induces a Riemannian metric on the space of positive definite matrices. The resulting manifold is called the Bures-Wasserstein space (after Bures, 1969); see Bhatia, Jain, and Lim (2019) for a comprehensive survey.

5.1 Gaussian Transport Splines

It is known that the Monge map from Gaussian $\mathcal{N}(0, \Sigma_1)$ to $\mathcal{N}(0, \Sigma_2)$ is the linear map T given by

$$T(X) = \Sigma_1^{-1/2} \big(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2} \big)^{1/2} \Sigma_1^{-1/2} X \qquad (6)$$

Cubic splines have the property that the interpolation evaluated at time t is a linear function of the interpolated points $(x_{t_i})_{i=0}^N$. That is, there is a linear map S_t (indexed by time) such that $t \mapsto S_t(x_{t_0}, \ldots, x_{t_N})$ is the cubic spline interpolating the data.[4]

This fact follows from the discussion in Appendix D and it has important consequences for our algorithm:

1. It implies that our algorithm outputs a process (X_t) such that X_t is a linear function of $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$. On the other hand, each X_{t_i} is a linear function of X_{t_0}, which follows from the description of Step 1 of our algorithm and the fact that Monge maps between Gaussians are linear (6).

Since a linear function of a Gaussian is also Gaussian, we conclude that *the transport spline interpolating Gaussian measures only passes through Gaussian measures.*

2. From the previous point, it is clear that the covariance matrix of X_t can be computed in terms of S_t, Σ_{t_0}, and the Monge maps (which have the closed-form expression (6)). We conclude that in this setting, not only can we output sample trajectories as in Algorithm 1, but *we can also efficiently output the covariance matrices of the interpolated measures.*

Furthermore, this discussion extends to any other interpolation method with this linearity property, such as higher-order splines, polynomial interpolation, and thin-plate splines.

We also remark that in the case where the data consists of *one-dimensional* Gaussian distributions, then in many cases the transport spline and the E-spline (described in Section 3) coincide.

Proposition 3. *Suppose that $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$ are one-dimensional Gaussians. Then, if the transport spline (μ_t) interpolating these data is never degenerate, i.e., μ_t is a non-degenerate Gaussian for each $t \in [0, 1]$, then the conditions of Theorem 1 hold.*

As discussed above, the transport spline through Gaussians automatically remains Gaussian, so the only hypothesis to check in this proposition is the nondegeneracy. See Appendix B for a discussion.

5.2 Approximation Guarantees

Our method is the first to provide approximation guarantees on Wasserstein space. In order to obtain strong quantitative results, we focus on the Bures-Wasserstein setting detailed in the previous section, where all measures $\mu_{t_i}^\star$ are centered non-degenerate Gaussian distributions.

The Bures-Wasserstein space has already been used in works such as Modin (2017) and Chewi et al. (2020b) as a prototypical setting in which to understand the behavior of algorithms set on the general Wasserstein space. Although the Bures-Wasserstein space is a Riemannian manifold and transport splines can in principle be studied using purely Riemannian techniques, we give proofs inspired by optimal transport so that the analysis may be more easily extended to other settings of interest.

We now state our main approximation result.

Theorem 2. *Let $(\mu_t^\star)$ be a curve of measures in Bures-Wasserstein space, and let $(X_t^\star) \sim (\mu_t^\star)$ be the Lagrangian coupling. Let:*

---

[3]The discussion here extends easily to incorporate non-centered measures. Indeed, the entire measure trajectory can be shifted by the Euclidean spline through the means μ_{t_i} to decouple the mean behavior; the property of mean-equivariance is shared by P-splines and transport splines, and likely with E-splines as well.

[4]Note that the matrix S_t is independent of $(x_{t_i})_{i=0}^N$, but depends on the time grid $(t_i)_{i=0}^N$.

- $L := \sup_{t \in [0,1]} \|\dot{X}_t^\star\|_{L^2(\mathbb{P})}$ *be the Lipschitz constant of the curve, and*

- $R := \sup_{t \in [0,1]} \|\ddot{X}_t^\star\|_{L^2(\mathbb{P})}$ *be an upper bound on its curvature, and*

- $\lambda_{\min}$ *be a lower bound on the eigenvalues of the covariance matrices of* $\mu_{t_0}^\star, \mu_{t_1}^\star, \ldots, \mu_{t_N}^\star$.

Let (μ_t) *be the cubic transport spline interpolating* $\mu_{t_0}^\star, \ldots, \mu_{t_N}^\star$ *and assume*

$$\alpha\delta \le t_i - t_{i-1} \le \delta, \qquad for\ i = 1, \ldots, N, \quad (7)$$

where $\alpha, \delta > 0$. *Then, provided that* $\delta < \sqrt{\lambda_{\min}}/(2L)$, *we have the following approximation guarantee:*

$$\sup_{t \in [0,1]} W_2(\mu_t, \mu_t^\star) \le \frac{58}{\alpha^3} R\delta^2.$$

The proof is given in Appendix C.

Some remarks:

1. The definition of L in the theorem agrees with the Lipschitz constant of $(\mu_t^\star)$ in the metric sense, as can be seen from Ambrosio, Gigli, and Savaré (2008, Theorem 8.3.1).

2. The quantity $\lambda_{\min}^{-1}$ can be interpreted as a bound on the curvature of Bures-Wasserstein space at the interpolation points; see Massart, Hendrickx, and Absil (2019) for details.

3. The $O(\delta^2)$ rate of convergence is optimal given our assumptions: a bound R on the second covariant derivative of the curve $(\mu_t^\star)$. Indeed, this matches classical approximation results for cubic splines on Euclidean space (Birkhoff and de Boor, 1964). We remark that under these assumptions, piecewise geodesic interpolation, where trajectories are piecewise linear and not differentiable, also achieves the $O(\delta^2)$ rate, and we give the proof of this in Appendix C.5. Of course, despite achieving the optimal rate in this class of curves, such interpolation is unsuitable for many applications (especially ones in which interpretation and visualization are a priority; see Figure 1).

4. We did not attempt to optimize the constant factor in Theorem 2 and it appears that it can, in fact, be improved; c.f. Remark 3

5. Cubic splines achieve higher-order approximation rates in the Euclidean setting, albeit over a restricted class of curves. For approximation of functions $f \in \mathcal{C}^k, k \le 4$, cubic splines enjoy a $O(\delta^k)$ approximation rate with explicit dependence on $\|f^{(k)}\|_{\sup}$. It is then natural to ask

whether it is possible to obtain rates better than $O(\delta^2)$ through a variant of transport splines. This can indeed be done by using more accurate approximations to the velocity vector fields (v_t); this study will be reported in a forthcoming work.

6 THIN-PLATE SPLINES

To demonstrate the flexibility of our method, we use transport splines to define a class of smooth interpolating surfaces on Wasserstein space. We first recall classical thin-plate splines. For a more complete account see Wahba (1990).

Thin-plate splines are the surface analog of cubic splines, and are useful in spatial problems where measurements are taken on a plane. Here, the times t_i are replaced with points $x_i \in \mathbb{R}^2$ at which we observe real values z_i. To account for this additional dimension the energy functional $\int_0^1 \|\ddot{\gamma}_t\|^2 \, dt$ that appears in the variational definition (3) of cubic splines is replaced by its bivariate counterpart. Thin-plate splines are defined as parametrized surfaces f that solve

$$\inf_f \int_{\mathbb{R}^2} \|\nabla^2 f\|_F^2 \ \text{ s.t. } \begin{cases} f : \mathbb{R}^2 \to \mathbb{R} \\ f(x_i) = z_i, \ i = 0, \ldots, N \end{cases} \quad (8)$$

where $\nabla^2 f$ is the Hessian of f, $\|\cdot\|_F$ denotes the Frobenius norm, and the interpolation data $(x_i, z_i) \in \mathbb{R}^2 \times \mathbb{R}$ is given. (Just as before, f is constrained to be $\mathcal{C}^2$.) It can be shown that (8) has a unique solution given by

$$f(x) = c_0 + c_1 x^{(1)} + c_2 x^{(2)} + \sum_{i=0}^N \alpha_i \phi(\|x - x_i\|)$$

where we use $x^{(i)}$ to denote coordinates, and

$$\phi(r) = r^2 \log r.$$

This leads to a closed form for the coefficients as follows. Let $K = (\phi(\|x_i - x_j\|))_{i,j=0}^N$ be the "kernel matrix" of the data, and define $P \in \mathbb{R}^{(N+1) \times 3}$ to have ith row $(1, x_i^{(1)}, x_i^{(2)})$.[5] Then let $L \in \mathbb{R}^{(N+4) \times (N+4)}$ be

$$L = \begin{bmatrix} K & P \\ P^\top & 0_{3 \times 3} \end{bmatrix}.$$

Letting $b = (z_0, \ldots, z_N, 0, 0, 0)$ be the padded data and $w = (\alpha_0, \ldots, \alpha_N, c_0, c_1, c_2)$ the coefficients from (6), these solve $Lw = b$. This can be inverted explicitly using the Schur complement, and in particular the resulting coefficients are linear in the data $(z_i)_{i=0}^N$.

---

[5]The function ϕ plays the role of a kernel for the reproducing kernel Hilbert space of twice-differentiable, finite-curvature surfaces, but it is *not* a kernel because it is not positive definite.

We now consider the measure-valued analog of the interpolation problem, namely, at each point x_i we observe a measure $\mu_{x_i}^\star$ and our goal is to find a smooth interpolating surface $x \mapsto \mu_x$ of measures.

As in the definition of E-splines, (8) can be generalized to Wasserstein space, but it is intractable for the same reasons. In contrast, applying Algorithm 1 is straightforward. Step 2 simply requires the fitting of a Euclidean thin-plate spline. For Step 1 we need only produce couplings between the observed measures $\mu_{x_i}^\star$.

One possiblity is to mimic the sequential coupling technique described in Section 4.1, namely we fix the ordering $x_0, x_1, \ldots, x_N$ and use the system of Monge maps $T_{i-1,i}$ taking $\mu_{x_{i-1}}^\star$ to $\mu_{x_i}^\star$. As before, we can draw $X_{x_0} \sim \mu_{x_0}^\star$ and then successively compute the random variables $X_{x_i} = T_{i-1,i}(X_{x_{i-1}}) \sim \mu_{x_i}^\star$ for all i. Sequential coupling is unsuitable here, however, because it distorts the geometry of the plane. To circumvent this issue, we next turn towards the special case when the measures $\mu_{x_i^\star}$ are defined over $\mathbb{R}$, which is already interesting enough to capture a breadth of applications.

The study of $\mathcal{P}_2(\mathbb{R})$ is greatly simplified by the fact that it is isometric to a convex subset of a Hilbert space and is therefore *flat*. Indeed, the special structure of $\mathcal{P}_2(\mathbb{R})$ has already been used fruitfully in many prior applications of optimal transport, such as curve registration (Panaretos and Zemel, 2016), geodesic principal components (Bigot et al., 2017), estimation of barycenters (Bigot et al., 2018), and uncoupled isotonic regression (Rigollet and Weed, 2019).

For our purposes, we will use the following key property of $\mathcal{P}_2(\mathbb{R})$: there is a unique coupling of all of the measures $\mu_{x_0}^\star, \mu_{x_1}^\star, \ldots, \mu_{x_N}^\star$ which is *simultaneously* optimal for every pair of measures. In other words, there exist random variables $X_{x_0}, X_{x_1}, \ldots, X_{x_N}$ such that for any $i, j = 0, 1, \ldots, N$, we have $X_{x_j} = T_{i,j}(X_{x_i})$, where $T_{i,j}$ is the Monge map from $\mu_{x_i}^\star$ to $\mu_{x_j}^\star$. Sampling from this coupling can be done using either of the of the following equivalent procedures:

1. Draw $X_{x_0} \sim \mu_{x_0}^\star$, and for each $i \in [N]$ let $X_{x_i} = T_{0,i}(X_{x_0})$ (the choice of x_0 does not affect the coupling).

2. Draw a uniform random variable U on $[0,1]$, and for $i = 0, 1, \ldots, N$ set $X_{x_i} = F_{\mu_{x_i}^\star}^{-1}(U)$, where F_μ denotes the CDF of μ.

See Appendix F.1 or Santambrogio (2015, §2.1-2.2).

In Figure 4 we display an application of thin-plate transport splines to temperature data. In the left-hand column we plot the quantiles of the interpolated measures. This is especially convenient when all of the
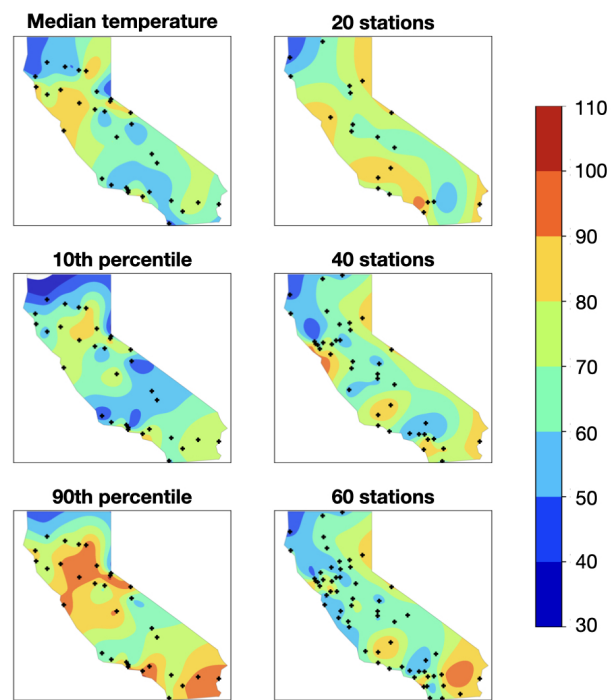

Figure 4: Thin-plate splines for California temperature data (in °F); in the left column are the quantiles, while in the right are the means of the interpolated measures for an increasing sample of observations. See Appendix F.3.

measures are Gaussian, in which case there is a simple and efficient algorithm for computing these quantiles (see Appendix F.2). The details for the experiment are given in Appendix F.3.

We conclude this section with a few remarks about the case of higher-dimensional measures, in which case there is no simultaneous optimal coupling of the measures. If we wish to use Monge map couplings as in Algorithm 1, one possibility is to first construct a tree graph whose vertices are the data $\mu_{x_i}^\star$, and use Monge map couplings along the edges of the tree. Here, the tree should be chosen to adequately capture the two-dimensional geometry of the spatial covariates. This consideration becomes especially relevant when the spatial covariates are sampled from a manifold, and it is of interest to combine our methodology with existing results on approximation of manifolds via graphs (Singer, 2006).

7 OPEN QUESTIONS

We conclude by discussing some interesting directions left open in this work. A natural question is to develop a computationally tractable notion of *smoothing* splines, and to investigate its statistical properties in the context of Wasserstein regression where the

$\mu_{t_i}^\star$ are observed with noise. As a second question, we remark that an approximation guarantee such as Theorem 2 can be compared with quantitative stability results for Monge maps (Gigli, 2011; Hütter and Rigollet, 2019) and extending such results to general Wasserstein space will likely require new techniques.

Acknowledgements

References

Altschuler, Jason, Weed, Jonathan, and Rigollet, Philippe (2017). "Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pp. 1961–1971.

Ambrosio, Luigi, Gigli, Nicola, and Savaré, Giuseppe (2008). *Gradient flows in metric spaces and in the space of probability measures*. Second. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, pp. x+334.

Backhoff-Veraguas, Julio et al. (2018). "Bayesian learning with Wasserstein barycenters". In: *arXiv e-prints*, arXiv:1805.10833.

Belkin, Mikhail et al. (2019). "Reconciling modern machine-learning practice and the classical bias–variance trade-off". In: *Proceedings of the National Academy of Sciences* 116.32, pp. 15849–15854.

Benamou, Jean-David, Gallouët, Thomas O., and Vialard, François-Xavier (2019). "Second-order models for optimal transport and cubic splines on the Wasserstein space". In: *Found. Comput. Math.* 19.5, pp. 1113–1143.

Bernton, Espen (2018). "Langevin Monte Carlo and JKO splitting". In: ed. by Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet. Vol. 75. Proceedings of Machine Learning Research. PMLR, pp. 1777–1798.

Bhatia, Rajendra, Jain, Tanvi, and Lim, Yongdo (2019). "On the Bures-Wasserstein distance between positive definite matrices". In: *Expo. Math.* 37.2, pp. 165–191.

Bigot, Jérémie et al. (2017). "Geodesic PCA in the Wasserstein space by convex PCA". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 53.1, pp. 1–26.

– (2018). "Upper and lower risk bounds for estimating the Wasserstein barycenter of random measures on the real line". In: *Electron. J. Stat.* 12.2, pp. 2253–2289.

Birkhoff, Garrett and de Boor, Carl (1964). "Error bounds for spline interpolation". In: *J. Math. Mech.* 13, pp. 827–835.

Bures, Donald (1969). "An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras". In: *Trans. Amer. Math. Soc.* 135, pp. 199–212.

Carrillo, J. A. and Vaes, U. (2019). "Wasserstein stability estimates for covariance-preconditioned Fokker-Planck equations". In: *arXiv e-prints*, arXiv:1910.07555.

Chen, Yongxin, Conforti, Giovanni, and Georgiou, Tryphon T. (2018). "Measure-valued spline curves: an optimal transport viewpoint". In: *SIAM J. Math. Anal.* 50.6, pp. 5947–5968.

Chewi, Sinho et al. (2020a). "Exponential ergodicity of mirror-Langevin diffusions". In: *NeurIPS*.

Chewi, Sinho et al. (2020b). "Gradient descent algorithms for Bures-Wasserstein barycenters". In: *Proceedings of Thirty Third Conference on Learning Theory*. Ed. by Jacob Abernethy and Shivani Agarwal. Vol. 125. Proceedings of Machine Learning Research. PMLR, pp. 1276–1304.

Chewi, Sinho et al. (2020c). "SVGD as a kernelized Wasserstein gradient flow of the chi-squared divergence". In: *NeurIPS*.

Cuturi, Marco (2013). "Sinkhorn distances: lightspeed computation of optimal transport". In: *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges et al. Curran Associates, Inc., pp. 2292–2300.

Durmus, Alain, Majewski, Szymon, and Miasojedow, Błażej (2019). "Analysis of Langevin Monte Carlo via convex optimization". In: *J. Mach. Learn. Res.* 20, Paper No. 73, 46.

Gigli, Nicola (2011). "On Hölder continuity-in-time of the optimal transport map towards measures along a curve". In: *Proc. Edinb. Math. Soc. (2)* 54.2, pp. 401–409.

– (2012). "Second order analysis on $(\mathscr{P}_2(M), W_2)$". In: *Mem. Amer. Math. Soc.* 216.1018, pp. xii+154.

Gousenbourger, Pierre-Yves, Massart, Estelle, and Absil, P.-A. (2019). "Data fitting on manifolds with composite Bézier-like curves and blended cubic

splines". In: *J. Math. Imaging Vision* 61.5, pp. 645–671.

Hütter, Jan-Christian and Rigollet, Philippe (2019). "Minimax rates of estimation for smooth optimal transport maps". In: *Ann. Statist. (to appear)*.

Jordan, Richard, Kinderlehrer, David, and Otto, Felix (1998). "The variational formulation of the Fokker-Planck equation". In: *SIAM J. Math. Anal.* 29.1, pp. 1–17.

Lu, Yulong, Lu, Jianfeng, and Nolen, James (2019). "Accelerating Langevin sampling with birth-death". In: *arXiv e-prints*, arXiv:1905.09863.

Massart, Estelle, Hendrickx, Julien M., and Absil, P.-A. (2019). "Curvature of the manifold of fixed-rank positive-semidefinite matrices endowed with the Bures-Wasserstein metric". In: *Geometric science of information*. Vol. 11712. Lecture Notes in Comput. Sci. Springer, Cham, pp. 739–748.

Modin, Klas (2017). "Geometry of matrix decompositions seen through optimal transport and information geometry". In: *J. Geom. Mech.* 9.3, pp. 335–390.

Noakes, Lyle, Heinzinger, Greg, and Paden, Brad (1989). "Cubic splines on curved spaces". In: *IMA Journal of Mathematical Control and Information* 6.4, pp. 465–473.

Panaretos, Victor M. and Zemel, Yoav (2016). "Amplitude and phase variation of point processes". In: *Ann. Statist.* 44.2, pp. 771–812.

Peyré, Gabriel and Cuturi, Marco (2019). "Computational optimal transport". In: *Foundations and Trends® in Machine Learning* 11.5-6, pp. 355–607.

Regev, Aviv et al. (2017). "Science Forum: The Human Cell Atlas". In: *eLife* 6. Ed. by Thomas R Gingeras, e27041.

Rigollet, Philippe and Weed, Jonathan (2019). "Uncoupled isotonic regression via minimum Wasserstein deconvolution". In: *Inf. Inference* 8.4, pp. 691–717.

Santambrogio, Filippo (2015). *Optimal transport for applied mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and their Applications. Calculus of variations, PDEs, and modeling. Birkhäuser/Springer, Cham, pp. xxvii+353.

Schiebinger, Geoffrey et al. (2019). "Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming". In: *Cell* 176.4, 928–943.e22.

Singer, Amit (2006). "From graph to manifold Laplacian: the convergence rate". In: *Appl. Comput. Harmon. Anal.* 21.1, pp. 128–134.

Villani, Cédric (2003). *Topics in optimal transportation*. Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, pp. xvi+370.

– (2009). *Optimal transport*. Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, pp. xxii+973.

Wahba, Grace (1990). *Spline models for observational data*. Society for Industrial and Applied Mathematics.

Zemel, Yoav and Panaretos, Victor M. (2019). "Fréchet means and Procrustes analysis in Wasserstein space". In: *Bernoulli* 25.2, pp. 932–976.