Fast and Smooth Interpolation on Wasserstein Space

Sinho Chewi  
schewi@mit.edu
Julien Clancy  
julienc@mit.edu
Thibaut Le Gouic  
tlegouic@mit.edu
Philippe Rigollet  
rigollet@mit.edu
George Stepaniants  
gstepan@mit.edu
Austin J. Stromme  
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 $\mu_t^*$ 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 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 \( C^2 \) sample paths and marginal laws \( X_t^\star \sim \mu_t^\star, t \in [0,1] \). Given \( \mu_t^0, \mu_t^1, \ldots, \mu_t^N \) at times \( 0 = t_0 < t_1 < \cdots < t_N = 1 \), the task is to construct a stochastic process \( (X_t)_{t \in [0,1]} \) such that \( X_t \) has \( C^2 \) sample paths and the distribution \( \mu_t \) of \( X_t \) interpolates the given measures, meaning \( \mu_t^i = \mu_t^\star \) 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 \(\mu^*_t\). In comparison, the method of Chen, Conforti, and Geor-
giou (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 $(\mu_t)$ 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 $\mu_0$, $\mu_1$ on $\mathbb{R}^d$ with finite second moment, the 2-Wasserstein distance $W_2$ is defined as

$$W_2^2(\mu_0, \mu_1) := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \|x - y\|^2 \, d\pi(x, y), \quad (2.1)$$

where $\Pi(\mu_0, \mu_1)$ is the set of all joint distributions with marginals $\mu_0$ and $\mu_1$. 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 $\mu_0$ has a density with respect to Lebesgue measure, then the solution of (2.1) 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 $\mu_0$ via $T$ is $\mu_1$ 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 (2.1) 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 $(\mu_t)$, 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 $\mu_t$ 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

$$v_t = \lim_{h \to 0} \frac{T_{\mu_t \to \mu_{t+h}} - \text{id}}{h} \quad \text{in } L^2(\mu_t) \quad (2.2)$$
where $T_{\mu_t \rightarrow \mu_{t+h}}$ is the Monge map between $\mu_t$ and $\mu_{t+h}$. For a proof see Ambrosio, Gigli, and Savaré (2008, Proposition 8.4.6).

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 $\mu_t$ 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

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

Choosing 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 $(\mu_t)$; 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

$$\min_{(\gamma_t)} \int_0^1 \|\dddot{\gamma}_t\|^2 \, dt \quad \text{s.t.} \quad \gamma_{t_i} = x_i \text{ for all } i.$$ (3.1)

The solution to this minimization problem is a piece-wise cubic polynomial that is globally $C^2$ 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 (3.1) the acceleration $\dddot{\gamma}$ is replaced with its Riemannian analogue, the covariant derivative $\nabla \dddot{\gamma}$ 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 $\nabla$. Thus, in analogy with the Riemannian setting, we can define energy splines (E-splines in short) via

$$\inf_{(\mu_t, v_t)} \int_0^1 \|\nabla v_t\|^2_{L^2(\mu_t)} \, dt \quad \text{s.t.} \quad \mu_{t_i} = \mu^*_{t_i} \text{ for all } i$$ (3.2)

where the minimization is taken over all curves $(\mu_t)$ and their tangent vectors $(v_t)$ (see Section 2). The solution to this problem naturally yields a stochastic process $(X_t)$ with marginal laws $(\mu_t)$, 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):

$$\inf_{(X_t)} \int_0^1 \mathbb{E}[\|\dot{X}_t\|^2] \, dt,$$ (3.3)
where the infimum is taken over stochastic processes \((X_t)\) with values in \(\mathbb{R}^d\) and such that \(X_{t_i} \sim \mu^*_{t_i}\) 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 \((3.3)\) can be reduced to a multimarginal optimal transport problem with quadratic cost.

Unfortunately, though solvable in principle, the formulation \((3.3)\) 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 non-degenerate Gaussian data \(\mu^*_{t_0}, \mu^*_{t_1}, \ldots, \mu^*_{t_N}\) such that there is a unique jointly Gaussian solution to the P-spline problem \((3.3)\) and it is not induced by a deterministic map.*

**Proposition 2** (informal). *There exist non-degenerate Gaussian data \(\mu^*_{t_0}, \mu^*_{t_1}, \ldots, \mu^*_{t_N}\) for which the E-spline \((3.2)\) and P-spline \((3.3)\) 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 the specified marginal laws \(\mu^*_{t_0}, \mu^*_{t_1}, \ldots, \mu^*_{t_N}\).
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 \(X_{t_0} \sim \mu^*_{t_0}\), 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 \(\mu^*_{t_{i-1}}\) to \(\mu^*_{t_i}\). 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 \(T_i \circ \cdots \circ T_1(X_{t_0}) \sim \mu^*_{t_i}\).

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((\(t_i\))\(i=0\)\(^N\), (\(\mu_{\star t_i}\))\(i=0\)\(^N\))
2: Draw \(X_{t_0}\) \(\sim\) \(\mu_{\star t_0}\)
3: for \(i = 1, \ldots, N\) do
4: \(X_{t_i} = T_i(X_{t_{i-1}})\), where \(T_i\) is the Monge map from \(\mu_{\star t_{i-1}}\) to \(\mu_{\star t_i}\)
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 \(\mu_{t_0}, \ldots, \mu_{t_N}\) sit along a curve of measures \((\mu_{t_i}^\star)\), then (as discussed in Section 2) there is an associated Lagrangian coupling \((X_{t_i}^\star)\) satisfying \(\dot{X}_{t_i}^\star = v_{t_i}^\star(X_{t_i}^\star)\). 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 (2.2) the Monge map \(T_1\) gives a first-order approximation to \(v_{t_0}^\star; T_1 - \text{id} = \delta v_{t_0}^\star + o(\delta)\) (see Ambrosio, Gigli, and Savaré (2008, Proposition 8.4.6)). Combining these approximations we get \(T_1(X_{t_0}^\star) = X_{t_1}^\star + o(\delta)\). 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 $\mu_{t_0}^*, \mu_{t_1}^*, \ldots, \mu_{t_N}^*$ are all one-dimensional, and for a measure $\mu$ let $F_{\mu}^t$ denote its quantile function.\footnote{Under our assumption that the measures are absolutely continuous, the quantile function $F_{\mu}^t$ simply coincides with the inverse CDF $F_{\mu}^{-1}$, but we use the quantile function notation here to reflect the general embedding $\mathcal{P}_2(\mathbb{R}) \hookrightarrow L^2[0, 1]$.} Let $(G_t)$ be the natural cubic spline in $L^2[0, 1]$ interpolating the quantile functions $F_{\mu_{t_0}}, F_{\mu_{t_1}}, \ldots, F_{\mu_{t_N}}$. Then:

**Theorem 1.** Suppose that for all $t$, $G_t$ is a valid\footnote{A valid quantile function $G_t : [0, 1] \to \mathbb{R} \cup \{\pm\infty\}$ is increasing and right-continuous.} quantile function. Then the transport spline and the E-spline (3.2) both coincide with the curve $(\mu_t)$ where $\mu_t$ has quantile function $G_t$. Furthermore, if $(X_t)$ is the stochastic process associated with the transport spline and $(X_t^*)$ is the Lagrangian coupling for the E-spline, then $(X_t)$ and $(X_t^*)$ 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-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 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.\footnote{The discussion here extends easily to incorporate non-centered measures.} 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

$$T(X) = \Sigma_1^{-1/2}(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\Sigma_1^{-1/2}X$$  \hspace{1cm} (5.1)

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

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 (5.1).

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 \( S_t, \Sigma_{t_0} \), and the Monge maps (which have the closed-form expression (5.1)). 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 \( \mu^*_{t_0}, \mu^*_{t_1}, \ldots, \mu^*_{t_N} \) are one-dimensional Gaussians. Then, if the transport spline \((\mu_t)\) interpolating these data is never degenerate, i.e., \( \mu_t \) 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 non-degeneracy. 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 \( \mu^*_{t_i} \) 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 \((\mu^*_t)\) be a curve of measures in Bures-Wasserstein space, and let \((X^*_t) \sim (\mu^*_t)\) be the Lagrangian coupling. Let:

- \( L := \sup_{t \in [0, 1]} \| \dot{X}^*_t \|_{L^2(P)} \) be the Lipschitz constant of the curve, and
- \( R := \sup_{t \in [0, 1]} \| \ddot{X}^*_t \|_{L^2(P)} \) be an upper bound on its curvature, and
- \( \lambda_{\min} \) be a lower bound on the eigenvalues of the covariance matrices of \( \mu^*_{t_0}, \mu^*_{t_1}, \ldots, \mu^*_{t_N} \).

Let \((\mu_t)\) be the cubic transport spline interpolating \( \mu^*_{t_0}, \ldots, \mu^*_{t_N} \) and assume

\[
\alpha \delta \leq t_i - t_{i-1} \leq \delta, \quad \text{for } i = 1, \ldots, N, \tag{5.2}
\]

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

\[
\sup_{t \in [0, 1]} W_2(\mu_t, \mu^*_t) \leq \frac{58}{\alpha^3} R \delta^2.
\]

The proof is given in Appendix C.

Some remarks:
1. The definition of \( L \) in the theorem agrees with the Lipschitz constant of \((\mu^*_t)\) 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 \((\mu^*_t)\). 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 \( f \in C^k, k \leq 4 \), cubic splines enjoy a \( O(\delta^k) \) approximation rate with explicit dependence on \( \|f^{(k)}\|_{\text{sup}} \). 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.

\section{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 (3.1) of cubic splines is replaced by its bivariate counterpart. Thin-plate splines are defined as parametrized surfaces \( f \) that solve

\begin{equation}
\inf \int_{\mathbb{R}^2} \|\nabla^2 f\|^2_{F} \quad \text{s.t.} \quad \begin{cases} 
 f : \mathbb{R}^2 \to \mathbb{R} \\
 f(x_i) = z_i, \; i = 0, \ldots, N 
\end{cases}
\end{equation}

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 \( C^2 \).) It can be shown that (6.1) 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 $i$th row $(1, x_i^{(1)}, x_i^{(2)})$. Then let $L \in \mathbb{R}^{(N+4)\times(N+4)}$ be

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

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 $(z_i)_{i=0}^N$.

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

As in the definition of E-splines, (6.1) 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 $\mu_{x_i}$.  

One possibility 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 $\mu_{x_{i-1}}$ to $\mu_{x_i}$. As before, we can draw $X_{x_0} \sim \mu_{x_0}$ and then successively compute the random variables $X_{x_1} = T_{1-0}(X_{x_0}) \sim \mu_{x_1}$ 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 $\mu_{x_i}$ 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 $\mu_{x_0}^{x_0}, \mu_{x_1}^{x_1}, \ldots, \mu_{x_N}^{x_N}$ 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 $\mu_{x_i}^{x_i}$ to $\mu_{x_j}^{x_j}$. Sampling from this coupling can be done using either of the following equivalent procedures:

1. Draw $X_{x_0} \sim \mu_{x_0}^{x_0}$, 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 $X_{x_i} = F_{\mu_{x_i}}^{-1}(U)$, where $F_{\mu}$ denotes the CDF of $\mu$.

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 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.

---

The function $\phi$ 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.
Fig 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.
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 \( \mu^*_i \), 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 \( \mu^*_t \) 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.

Acknowledgments. Sinho Chewi and Austin J. Stromme were supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. Julien Clancy and George Stepaniants were supported by the NSF GRFP. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. Thibaut Le Gouic was supported by ONR grant N00014-17-1-2147 and NSF IIS-1838071. Philippe Rigollet was supported by NSF awards IIS-1838071, DMS-1712596, DMS-T1740751, and DMS-2022448.

APPENDIX A: DETAILS FOR THE P-SPLINE EXAMPLE

In this section we consider the following spline problem: for \( N > 1 \) and times \( t_i = i/N, i = 0, \ldots, N \), suppose we observe

\[
\mu^*_i := N(0, (1 - t_i)^2 + t_i^2), \quad i = 0, \ldots, N.
\]  

(A.1)

This is the data for which we make the claims in Propositions 1 and 2.

A.1 Proposition 1

We begin by remarking that in general, there is no reason to expect that solutions of the P-spline problem (3.3) are deterministic. Indeed, consider the following.

\textbf{Proposition 4.} Let \( \mu_0^* \) and \( \mu_1^* \) be any probability measures. Then, any coupling \((X_0, X_1)\) of the two measures induces an optimal P-spline solution \((X_t)\) to (3.3) with data \( \mu_0^* \) and \( \mu_1^* \).

\textbf{Proof.} Indeed, simply set \( X_t := (1 - t)X_0 + tX_1 \). Since \( t \mapsto X_t \) is a line traversed at constant speed, it incurs zero P-spline cost and is therefore optimal for (3.3). \( \square \)

As this example shows, the P-spline problem with two measures is quite degenerate; in particular, it does not recover the \( W_2 \) geodesic joining \( \mu_0 \) to \( \mu_1 \), and \( X_t^* \) is not guaranteed to be a deterministic function of \( X_0^* \). A slight modification of this simple example yields:
PROPOSITION 5. Let \( \mu_0^\ast \) be any absolutely continuous measure. Then, there exist absolutely continuous data \( (\mu_{i/N}^\ast)_{i=1}^N \) and an optimal solution \( (X_t) \) to the P-spline problem (3.3) for \( (\mu_{i/N}^\ast)_{i=0}^N \) such that \( X_1 \) is not a deterministic function of \( X_0 \).

PROOF. Indeed, let \( T, \tilde{T} : \mathbb{R}^d \to \mathbb{R}^d \) be two mappings which are \( \mu_0^\ast \)-a.e. distinct, i.e., \( T \neq \tilde{T} \). Draw \( X_0 \sim \mu_0^\ast \). Then, we either set \( X_t = (1 - t)X_0 + tT(X_0) \) or else \( X_t = (1 - t)X_0 + t\tilde{T}(X_0) \) with probability 1/2 each (with the choice being made independently of the draw of \( X_0 \)). Set \( \mu_{i/N}^\ast := \text{law}(X_{i/N}) \).

By construction, the marginals of the process \( (X_t) \) at times \( 0, 1/N, \ldots, 1 \) do indeed interpolate the data. Also, since \( t \mapsto X_t \) is a straight line traversed at constant speed, then \( (X_t) \) incurs zero P-spline cost and is optimal for (3.3).

Since \( T \) and \( \tilde{T} \) are distinct, \( X_1 \) is not a deterministic function of \( X_0 \). Also, the mappings \( T \) and \( \tilde{T} \) can easily be chosen to make the data all absolutely continuous (e.g., by taking them to be gradients of uniformly convex functions; c.f. the proof of Villani (2003, Proposition 5.9)). □

(Compare this with Proposition 7 and the subsequent remark in Benamou, Gallouët, and Vialard (2019).)

We next turn towards the Gaussian case. As detailed in Chen, Conforti, and Georgiou (2018) and Benamou, Gallouët, and Vialard (2019), the P-spline problem (3.3) can be reduced to a multimarginal optimal transport problem involving the measures \( \mu_{0}^\ast, \mu_{1}^\ast, \ldots, \mu_{t/N}^\ast \),

\[
\inf_{\pi \in \Pi(\mu_{0}^\ast, \mu_{1}^\ast, \ldots, \mu_{t/N}^\ast)} \int c \, d\pi, \tag{A.2}
\]

where \( c \) is a quadratic cost function. The reduction is in the following sense: if \( \pi \) is an optimal solution for (A.2), then let \( (X_{t_0}, X_{t_1}, \ldots, X_{t_N}) \sim \pi \), and fit a Euclidean cubic spline \( (X_t) \) through the points \( (X_{t_i})_{i=0}^N \). Then, the stochastic process \( (X_t) \) is an optimal solution for (3.3). Any optimal solution of (3.3) is also of this form, having sample paths that are cubic splines.

Since the cost in the multimarginal problem (A.2) is quadratic, it depends only on the mean and covariance matrix of the coupling \( \pi \). Suppose now that the data \( (\mu_{i/N}^\ast)_{i=0}^N \) is Gaussian, and suppose we are given any optimal coupling \( \pi \) for (A.2). Then, we can find a jointly Gaussian coupling \( \tilde{\pi} \) of the data which has the same mean and covariance structure as \( \pi \), which means \( \tilde{\pi} \) is also optimal for (3.3). The coupling \( \tilde{\pi} \) then induces a Gaussian process \( (\tilde{X}_t) \) which is optimal for (3.3). Such a solution has the appealing property that the law \( \mu_t \) of \( \tilde{X}_t \) is also Gaussian for every time \( t \).

From this discussion, it is natural to restrict ourselves to solutions to (3.3) which are Gaussian processes. We call such a solution a Gaussian solution to the P-spline problem (3.3). We now state a counterexample which proves Proposition 1.

PROPOSITION 6. Assume \( N > 1 \). For \( i = 0, 1, \ldots, N \), let \( \mu_{t_i}^\ast = \mathcal{N}(0, (1 - t_i)^2 + t_i^2) \). Then there is a unique Gaussian solution to the P-spline problem (3.3) and it is not induced by a deterministic map.

PROOF. The key observation is that the marginals \( \mu_{t_i}^\ast \) arise from the curve of measures formed as the law of \( X_i^\ast := (1 - t)X_i^\ast + tX_i^\ast \) for independent standard Gaussians \( X_i^\ast \) and \( X_i^\ast \). If we consider the distribution on paths which is the law of \( (X_t^\ast) \), then it is supported on straight lines traversed
at constant speed and so it must be optimal for the $P$-spline problem (3.3), having zero objective value.

Consider some other stochastic process $(X_t)$ such that the law of $(X_t)_{i=0}^N$ is jointly Gaussian. For $(X_t)$ to be an optimal solution to the $P$-spline problem (3.3), it must also have zero objective value and hence be supported on straight lines almost surely. Thus, we must have $X_t = (1-t)X_0 + tX_1$. By the marginal constraints we have $E[X_0] = E[X_1] = 1$ and so long as $N > 1$, for $i = 1, \ldots, N-1$, it holds that $t_i \notin \{0,1\}$ and

$$(1 - t_i)^2 + t_i^2 = E[(1 - t_i)X_0 + t_iX_1]^2$$

$$= (1 - t_i)^2 + t_i^2 + 2t_i (1 - t_i) E[X_0X_1].$$

Therefore $E[X_0X_1] = 0$ and $(X_t)$ has the same distribution as $(X_t^*)$. Consequently, the unique jointly Gaussian solution to the $P$-spline problem is $(X_t^*)$. Clearly, the path $(X_t^*)$ is not a deterministic function of $X_0$. Indeed, $X_1$ is independent of $X_0^*$.

**Remark 1.** The uniqueness assertion is false when $N = 1$, even when restricting to Gaussian solutions, which again highlights that the $P$-spline problem between two measures is degenerate.

**A.2 Proposition 2**

In this section we provide the proof of Proposition 2. Understanding E-splines requires a few technical results, which we first collect before moving on to the proof. We remark that, prior to this work, little was known about E-splines. In particular, it was not known whether the E-spline interpolation of Gaussian measures consists only of Gaussian measures.

Throughout, it will be convenient to consider the E-spline problem over the closed convex set of curves taking values in a closed convex set $K$ of a Hilbert space:

$$\min_{\gamma: [0,1] \to K} \int_0^1 \|\dot{\gamma}(t)\|^2 \, dt \quad \text{s.t.} \quad \gamma(t_i) = x_i \quad \text{for all } i$$

$$(E_K)$$

Denote by $E[\gamma] = \int_0^1 \|\dot{\gamma}(t)\|^2 \, dt$ the objective function in $(E_K)$. It follows from the triangle inequality and strict convexity of the function $x \mapsto x^2$ that $E$ is strictly convex on the convex set of admissible curves, so the solution must be unique if it exists. We denote this unique solution by $\gamma_K$.

**Proposition 7.** Let $H$ be a Hilbert space, and let $L \subseteq H$ be a closed linear subspace. Take points $x_0, \ldots, x_N \in L$. Then the solution $\gamma_H$ of the E-spline problem $(E_H)$ on $H$ satisfies $\gamma_H(t) = \gamma_L(t) \in L$ for all $t$.

**Proof.** Let $P$ be the orthogonal projection onto $L$, and suppose $\gamma$ interpolates the points $(x_i)_{i=0}^N$. Then for any admissible curve $\gamma(t) = P\gamma(t) + (I - P)\gamma(t)$, so $\dot{\gamma}(t) = P\dot{\gamma}(t) + (I - P)\dot{\gamma}(t)$ as well.

Since these two terms are orthogonal, we have

$$\|\dot{\gamma}(t)\|^2 = \|P\dot{\gamma}(t)\|^2 + \|(I - P)\dot{\gamma}(t)\|^2.$$

Thus, on the one hand, if $\dot{\gamma}(t) = P\gamma_H(t)$ then $E[\dot{\gamma}] \leq E[\gamma_H]$, and $\dot{\gamma}$ is interpolating because $x_i \in L$. On the other hand, $E[\gamma_H] \leq E[\gamma_L] \leq E[\dot{\gamma}]$ and by uniqueness, $\gamma_H = \gamma_L$. □
Proposition 8. Let $K$ be a convex subset of a Hilbert space $H$ whose span is closed, and let $x_1, \ldots, x_n \in K$. If $\gamma_K(t)$ lies in the relative interior of $K$ for all times $t$, then $\gamma_K = \gamma_H$.

Proof. Let $L$ be the linear span of $K$, which is closed. In light of Proposition 7, it suffices to prove that $\gamma_K = \gamma_L$ so replacing $H$ by $L$ we may assume that $K$ is of full dimension.

Let $f: [0,1] \to H$ be a twice differentiable perturbation such that $f(t_i) = 0$ for all $i$. Hence, $\gamma_K + \varepsilon f$ is admissible for $(E_H)$. Since $\gamma_K$ lies in the interior of $K$ and $K$ is full-dimensional, a standard compactness argument shows that for any such $f$ there exists an $\varepsilon > 0$ with $\gamma_K(t) + \varepsilon f(t) \in K$ for all $t$. By optimality of $\gamma_K$ we then have $E[\gamma_K + \varepsilon f] \geq E[\gamma_K]$. Thus $\gamma_K$ is stationary for $E$ considered on $H$, and because $E$ is strictly convex it follows that $\gamma_K$ is optimal for $(E_H)$ and is therefore equal to $\gamma_H$ by uniqueness. □

Proposition 9. Let $\mu_{\varepsilon_0}, \mu_{\varepsilon_1}, \ldots, \mu_{\varepsilon_N}$ be Gaussian measures on $\mathbb{R}$. Consider the Gaussian version of the E-spline problem on $\mathbb{R}$:

$$\min_{(\gamma_t)} \int_0^1 \|\nabla v_t\|^2_{L^2(\gamma_t)} \, dt \quad \text{s.t.} \quad \gamma_t = \mu_{\varepsilon_t}, \, t = 1, \ldots, N,$$

where the minimization is taken over curves $(\gamma_t)$ of Gaussian measures with their corresponding tangent vectors $(v_t)$ (as described in Section 2). That is, it is the Wasserstein E-spline problem (3.2) in $P_2(\mathbb{R})$ with the added constraint that the measures are Gaussian. If there is an optimal solution $(\gamma_{\varepsilon_t}^*)$ which is non-degenerate Gaussian for all time, then it is also the solution to the E-spline problem (3.2).

Proof. It is known that $P_2(\mathbb{R})$ is isometric to a closed convex subset $S$ of the Hilbert space $H = L^2[0,1]$ (see the discussion following Ambrosio, Gigli, and Savaré, 2008, Lemma 9.1.4). This isometry is given by $\mu \mapsto F^\dagger_\mu$, where $F^\dagger_\mu$ denotes the quantile function of $\mu$. Let $K$ be the image of the mean-zero Gaussian measures under this isometry; it is immediate that $K$ is convex, since the Gaussian measures form a geodesically convex set in $P_2(\mathbb{R})$, and it has closed span because it is finite-dimensional. In light of this isometry the E-spline problem (3.2) is equivalent to $(E_S)$ while the Gaussian E-spline problem stated in the proposition is equivalent to $(E_K)$ and $\gamma^* = \gamma_K$ (the preservation of E-splines under isometry is discussed in Appendix A.3).

Applying Proposition 8 to $\gamma^* = \gamma_K$, we deduce that $\gamma^* = \gamma_H$. Moreover, $E[\gamma_H] \leq E[\gamma_S] \leq E[\gamma^*]$, whence by uniqueness we get that $\gamma^* = \gamma_S$ as well. □

We also require a technical lemma regarding P-splines which remain Gaussian for all times, which follows from considerations of several-variable complex functions.

Lemma 1. Let $(\mu_t)$ be a P-spline with initial and final data $\mu_0$ and $\mu_1$ which are Gaussian, and assume:

1. $\mu_t$ is a Gaussian distribution for all times $t$,
2. $(\mu_t)$ has zero cost for the P-spline objective.

Then $(\mu_t)$ is induced by a jointly Gaussian coupling of $\mu_0$ and $\mu_1$. 

Proof. Since \((\mu_t)\) has zero cost it must be supported on straight lines, so if we let \(X_t \sim \mu_t\) where these are coupled according to the \((\mu_t)\) coupling, then
\[
X_t = (1 - t)X_0 + tX_1
\]
(A.3)
and by assumption this variable is Gaussian. Let \(Z\) be the Gaussian with the same covariance structure as \(X\). Scaling (A.3) by a positive constant, we get, for all \(a, b \geq 0\)
\[
\langle (a, b), X \rangle = \langle (a, b), Z \rangle
\]
where we mean equality in distribution. This implies
\[
\varphi_X(a, b) = \varphi_Z(a, b)
\]
where \(\varphi_Y\) denotes the characteristic function of \(Y\) and is defined by \(\varphi_Y(z) = \mathbb{E}[e^{i(z,Y)}]\). Now, it is well-known that if \(\mathbb{E}[e^{m\|Y\|}] < \infty\) for some \(m > 0\) then \(\varphi_Y\) continues to a holomorphic function in the strip \(\{z \mid |\text{Im } z| < m \text{ for } i = 1, \ldots, d\}\) (Lehmann and Romano, 2005, Theorem 2.7.1). In particular, if \(Y\) has sub-Gaussian tails, \(\varphi_Y\) is entire.

Functions of several complex variables admit an identity theorem, similar to the univariate complex case, which can be found in Range (1986, Remark 1.20).\(^6\) This is:

**Theorem** (identity theorem). Let \(f\) and \(g\) be holomorphic functions of several complex variables in a domain \(\Omega \subseteq \mathbb{C}^d\), and let \(z \in \Omega\). A real cube of radius \(r\) about \(z\) is defined as
\[
\{(z_1 + x_1, \ldots, z_d + x_d) \in \mathbb{C}^d \mid |\text{Re } x_i| < r \text{ for } i = 1, \ldots, d\}.
\]
If \(f\) and \(g\) agree on a real cube of positive radius about \(z\), then \(f \equiv g\) on all of \(\Omega\).

Now, \(X\) has sub-Gaussian tails. Indeed,
\[
M_X(t) = \mathbb{E}e^{(t,X)} = \mathbb{E}e^{t_1X_0 + t_2X_1} \leq \left(\mathbb{E}e^{2t_1X_0} \mathbb{E}e^{2t_2X_1}\right)^{1/2} = e^{t_1^2 \text{var } X_0 + t_2^2 \text{var } X_1}
\]
where \(M_X\) denotes the moment generating function of \(X\). Thus \(\varphi_X\) is entire, along with \(\varphi_Z\), and it is clear from the above discussion that they agree on the real cube about \(z = (1, 1)\) with radius \(r = 1\). The identity theorem then implies that \(\varphi_X \equiv \varphi_Z\), so \(X \overset{d}{=} Z\). Thus \(X\) is jointly Gaussian. \(\square\)

Proposition 2 is implied by the following result.

**Proposition 10.** For \(i = 0, \ldots, N\), let \(\mu^*_t = N(0, \sigma^2_t)\), where \(\sigma^2_t = (1 - t)^2 + t^2\). Then for all \(N \geq 2\), the \(E\)-spline (3.2) and \(P\)-spline (3.3) interpolations do not coincide.

\(^6\)The careful reader will note that the hypothesis of this theorem is much stronger than the single-variable requirement that \(f\) and \(g\) agree merely on a set with an accumulation point. For several complex variables this is not sufficient; indeed, several-variable holomorphic functions never have isolated zeros.
Before starting the proof, we dispense with a possible source of confusion. The solution to the P-spline problem (3.3) is a stochastic process \((X_t)\); on the other hand, the E-spline solution yields a natural stochastic process, namely the Lagrangian coupling \((X^*_t)\) (see Section 2). In the proposition, we are not asserting that the process \((X_t)\) and \((X^*_t)\) are different (indeed this is an easier statement to prove since the P-spline solution is often not even deterministic; see Appendix A.1). Instead, we are asserting that the interpolated measures associated with the E- and P-splines are different, which is a synonym for “rigorous.”

**Proof.** First, the manifold of mean-zero Gaussian measures on \(\mathbb{R}\) equipped with the \(W_2\) metric is isometric to the ray \([0, \infty)\) equipped with the standard Euclidean metric. Indeed, we have

\[
W_2(\mathcal{N}(0, \sigma_0^2), \mathcal{N}(0, \sigma_1^2)) = |\sigma_0 - \sigma_1|.
\]

Suppose we have data \(\mu^*_i = \mathcal{N}(0, \sigma_i^2)\) at times \(t_i\) and let \(t \mapsto \gamma(t)\) be the Euclidean spline interpolation of \((t_i, \sigma_i)\) on \(\mathbb{R}\). It is possible that \(\gamma(t) \leq 0\) at some \(t\), but if \(\gamma(t) > 0\) for all \(t\), then by Proposition 8 it must also be the spline considered on the ray \([0, \infty)\). Since covariant derivatives are preserved under isometry (see Appendix A.3 for a formal verification in our setting), the function \(E[\cdot]\) is also preserved under isometry, and so its minimizers — E-splines — are preserved as well. This means that the Gaussian-constrained E-spline is

\[
\mu^E_t = \mathcal{N}(0, \gamma(t)^2), \quad t \in [0, 1],
\]

and by Proposition 9 this must coincide with the Wasserstein E-spline (3.2). This is all under the hypothesis that \(\gamma(t) > 0\).

Now substitute our example, with \(\sigma_i^2 = (1 - t_i^2)^2 + t_i^2\). We need to check that \(\gamma(t)\) remains strictly positive for all times. From Hall and Meyer (1976, Theorem 5), we see that for all \(t\)

\[
|\gamma(t) - \sqrt{t^2 + (1 - t)^2}| \leq \frac{5}{384} \cdot 24\sqrt{2} \cdot \frac{1}{N^4}.
\]

For \(N \geq 2\) this is less than 0.03. The smallest value of \(\sqrt{t^2 + (1 - t)^2}\) is \(\sqrt{1/2} \approx 0.7071\), so the spline is bounded below by 0.704 for all times.

Let \((\mu^P_t)\) be an interpolating P-spline. It is possible that this is not unique, but if \(\mu^P_t\) is not Gaussian for some \(t\) then we are done, since \(\mu^E_t\) is Gaussian by Proposition 9. Applying Lemma 1, we see that \(\mu^P_t\) must be induced by a jointly Gaussian coupling of \(\mu^P_0\) and \(\mu^P_1\), so by Proposition 6 it must be that \(\mu^P_t = \mathcal{N}(0, (1 - t)^2 + t^2)\).

The standard deviation of \(\mu^E_t\) is \(\gamma(t)\) and this is locally a cubic polynomial in \(t\). The standard deviation of the P-spline \(\mu^P_t\), however, is given by \(\sqrt{(1 - t)^2 + t^2}\), which cannot be locally represented by a polynomial, so they must differ. \(\Box\)

From the final steps of our proof, we see that (in the Gaussian case) P-splines and E-splines will most likely differ generically, since their interpolated variances are polynomial splines of different orders.

**A.3 Preservations of Splines under Isometry**

In this section, we give a formal\(^7\) verification of the assertion that the E-spline functional is preserved under the isometry between \(\mathcal{P}_2(\mathbb{R})\) and its image in \(H = L^2[0, 1]\). Formally, this assertion

\(^7\)The word *formal* here, meaning that the argument proceeds by manipulating the *form* of the expressions, is not a synonym for “rigorous”.

can be viewed as a manifestation of a classical fact from Riemannian geometry: the covariant derivative (associated with the Levi-Civita connection) depends only on the Riemannian metric, and is thus preserved under isometries.\textsuperscript{8}

In the derivation below, we make all necessary regularity assumptions (e.g., we can assume that the measures are compactly supported) in order to convey the intuition. Suppose \((\mu_t)\) is a curve of measures in \(\mathcal{P}_2(\mathbb{R})\) and let \((v_t)\) be the corresponding tangent vectors. The relationship between \((\mu_t)\) and \((v_t)\) is given by the \textit{continuity equation} (Ambrosio, Gigli, and Savaré, 2008, Theorem 8.3.1):

\[
\partial_t \mu_t + (\mu_t v_t)' = 0. 
\]  

(A.4)

Here, we use \(\partial_t\) for the time derivative, and we use \('\) to denote spatial derivatives. If \(F_\mu\) denotes the CDF of \(\mu\), then (A.4) implies

\[
\partial_t F_\mu(x) = \partial_t \int_{-\infty}^{x} \mu_\tau \, d\tau = - \int_{-\infty}^{x} (\mu_t v_t)' = -\mu_t(x) v_t(x).
\]

Next, if we differentiate the relation \(F_\mu^{-1}(F_\mu(x)) = x\), we obtain

\[
0 = (\partial_t F_\mu^{-1})(F_\mu(x)) + (F_\mu^{-1})'(F_\mu(x))
\]

\[
= (\partial_t F_\mu^{-1})(F_\mu(x)) + \frac{1}{F_\mu'(x)}
\]

\[
= (\partial_t F_\mu^{-1})(F_\mu(x)) + \frac{1}{\mu_t'(x)},
\]

where we have applied the inverse function theorem. Thus,

\[
(\partial_t F_\mu^{-1})(\alpha) = v_t(F_\mu^{-1}(\alpha)).
\]

(A.5)

Differentiating again,

\[
(\partial_t^2 F_\mu^{-1})(\alpha) = (\partial_t v_t)(F_\mu^{-1}(\alpha)) + (\partial_t^2 v_t)(F_\mu^{-1}(\alpha)) (\partial_t F_\mu^{-1})(\alpha)
\]

\[
= (\partial_t v_t + v_t' v_t)(F_\mu^{-1}(\alpha)).
\]

However, we recognize \(\partial_t v_t + v_t' v_t\) as the covariant derivative \(\nabla v_t v_t\) in \(\mathcal{P}_2(\mathbb{R})\) (see for example the discussion in Chen, Conforti, and Georgiou, 2018, §5.1). In particular, it implies

\[
\int_0^1 |\partial_t^2 F_\mu^{-1}|^2 = \int_0^1 |(\partial_t v_t + v_t' v_t)\circ F_\mu^{-1}|^2
\]

\[
= \int_0^1 |\partial_t v_t + v_t' v_t|^2 \, d\mu_t
\]

\[
= \|\nabla v_t v_t\|_{L^2(\mu_t)}^2,
\]

where we use the fact that the pushforward of the uniform distribution on \([0,1]\) under \(F_\mu^{-1}\) is \(\mu_t\). This equation shows that the norm (measured in \(H\)) of the acceleration of the curve \(t \mapsto F_\mu^{-1}\) in \(H\) is the same as the norm (measured in \(\mathcal{P}_2(\mathbb{R})\)) of the acceleration of the curve \(t \mapsto \mu_t\) in \(\mathcal{P}_2(\mathbb{R})\), and thus the E-spline cost functional is preserved by the embedding \(\mathcal{P}_2(\mathbb{R}) \hookrightarrow H\).

\textsuperscript{8}In fact, this is related to Gauss’s famous \textit{Theorema Egregium}, see Carmo (2016, §4.3) and Carmo (1992, Remark 2.7).
Remark 2. From the equation (A.5), we can also read off the isometry between the tangent space of $H$ and the tangent space of $P_2(\mathbb{R})$.

The reader who is uncomfortable with the formal derivation above can instead use the isometric embedding $P_2(\mathbb{R}) \hookrightarrow L^2[0,1]$ as the definition of the geometry of $P_2(\mathbb{R})$ (and thus, the definition of E-splines on $P_2(\mathbb{R})$). Indeed, a rigorous development of second-order calculus on Wasserstein space faces significant technical hurdles (Gigli, 2012), and such a definition is actually more convenient for the purposes of this paper.

APPENDIX B: E-SPLINES AND TRANSPORT SPLINES IN ONE DIMENSION

In this section, we investigate the relationship between transport splines and E-splines on $P_2(\mathbb{R})$, leading to a proof of Theorem 1. We will use the calculation in Appendix A.3, and moreover we recommend that readers read Appendix A before this section in order to gain familiarity with E-splines.

Recall also that we assume that the measures $\mu^c_t$ are absolutely continuous in order to properly define the covariant derivative. However, the embedding $P_2(\mathbb{R}) \hookrightarrow L^2[0,1]$ allows us to rigorously extend the definition of an E-spline on all of $P_2(\mathbb{R})$.

Proof of Theorem 1. Let $U$ be a uniform random variable on $[0,1]$, and define the random variables

$$X_t := F^\dagger_{\mu^c_t}(U) \sim \mu^c_t, \quad i = 0, 1, \ldots, N.$$ 

From the discussion in Appendix F.1, these random variables are simultaneously optimally coupled. In particular, each successive pair of these random variables is coupled via a Monge map. It follows from the definition of a transport spline that the stochastic process $(X_t)$ associated with the transport spline can be realized as the (Euclidean) cubic spline interpolating the points $(X_t)_{i=0}^N$.

Since each $X_t$ is a function of $U$, so is the interpolation $X_t$, so we can write $X_t = \tilde{G}_t(U)$. It follows that $(\tilde{G}_t)$ is the cubic spline in $H = L^2[0,1]$ which interpolates the quantiles $(F^\dagger_{\mu^c_t})_{i=0}^N$, that is, $(\tilde{G}_t) = (G_t)$. At this point, we have established one of the assertions of Theorem 1, namely, the explicit description of the process $(X_t)$ associated with the transport spline.

Next, since $X_t = G_t(U)$, by hypothesis $G_t$ is an increasing function that pushes forward the uniform distribution to the law $\mu_t$ of $X_t$. By the characterization of Monge maps in one dimension (Appendix F.1), it follows that $G_t = F^\dagger_{\mu_t}$. Since $(G_t)$ is a cubic spline, then it minimizes curvature, i.e., it solves the problem

$$\inf_{(G_t)} \int_0^1 ||\dddot{G}_t||_{L^2[0,1]}^2 \, dt, \quad \text{s.t.} \quad G_t = F^\dagger_{\mu_t} \text{ for all } i.$$ 

From our characterization $G_t = F^\dagger_{\mu_t}$, it is clear that $(\mu_t)$ solves the problem

$$\inf_{(\mu_t)} \int_0^1 ||\dot{\partial}_t^2 F^\dagger_{\mu_t}||_{L^2[0,1]}^2 \, dt, \quad \text{s.t.} \quad \mu_t = \mu^c_t \text{ for all } i,$$

since the first problem is a relaxation of the second (given a solution $(\mu_t)$ of the second problem, we can obtain a solution $(G_t) = (F^\dagger_{\mu_t})$ for the first problem). Indeed, the second problem can be
interpreted as the first problem with the additional constraint that the functions $G_t$ must be quantile functions. Next, in light of the isometry described in Appendix A.3, the latter problem is equivalent to

$$\inf_{(\mu_t, v_t)} \int_0^1 \|\nabla v_t \|_{L^2(\mu_t)}^2 \, dt, \quad \text{s.t.} \quad \mu_{ti} = \mu_{ti}^\star \text{ for all } i,$$

where the infimum is taken over curves $(\mu_t)$ in $P_2(\mathbb{R})$ and their corresponding tangent vectors $(v_t)$. This problem is seen to be the E-spline problem (3.2).

We have thus shown that $(\mu_t)$ is an E-spline. Actually, in light of Proposition 7 and the fact that $(G_t)$ is the spline in $H$, then the E-spline is unique. Thus, the E-spline and transport spline coincide.

Finally, it remains to show that the Lagrangian coupling $(X_t^c)$ associated with the E-spline has the same law as $(X_t)$. For this, we can simply appeal to the embedding $P_2(\mathbb{R}) \hookrightarrow H$ again. Indeed, since $\dot{X}_t = \partial_t P_{\mu_t}^\star(U)$, the calculation in Appendix A.3 shows that $\dot{X}_t = v_t(X_t)$ where $(v_t)$ is the tangent vector to $(\mu_t)$, so in fact $(X_t)$ is the Lagrangian coupling of $(\mu_t)$. □

In particular, since the Gaussian measures form a 2 dimensional half-subspace of $L^2[0,1]$ with the usual identification $P_2(\mathbb{R}) \hookrightarrow L^2[0,1]$, the E-spline interpolation between Gaussian measures is the transport spline if transport splines is not degenerate at any time (i.e., the transport lies in the relative interior of Gaussian measures within $P_2(\mathbb{R})$). This yields Proposition 3.

We conclude this section by giving some examples showing that E-splines and transport splines can differ when the spline $(G_t)$ described in Theorem 1 does not stay within $P_2(\mathbb{R}) \subset L^2[0,1]$. First, we give a simple Gaussian counterexample.

**Proposition 11.** Let $\delta > 0$ be sufficiently small and consider the measures

$$\mu_0^\star = \mu_1^\star = \mathcal{N}(0,1), \quad \mu_{1/3}^\star = \mu_{2/3}^\star = \mathcal{N}(0,\delta^2).$$

Then, the E-spline (3.2) interpolation $(\mu_t^E)$ and transport spline interpolation $(\mu_t^T)$ do not coincide for this data.

**Proof.** Let $(X_t)$ denote the stochastic process corresponding to the transport spline. It is easy to see that $(X_0, X_{1/3}, X_{2/3}, X_1) = (X_0, \delta X_0, \delta^2 X_0, X_0)$ is the optimal coupling at the knots. If we let $S_t$ denote the linear mapping which produces the spline (as introduced in Section 5), it follows that

$$X_t = S_t(X_0, \delta X_0, \delta^2 X_0, X_0) = S_t(1, \delta, \delta, 1)X_0,$$

so that $\mu_t^\star = \mathcal{N}(0, S_t(1, \delta, \delta, 1)^2)$.

If we identify the space of Gaussians with the half-ray $[0, \infty)$, then the transport spline corresponds to the curve of standard deviations $t \mapsto |S_t(1, \delta, \delta, 1)|$. However, because the spline curve $t \mapsto S_t(1,0,0,1)$ becomes negative between $1/3$ and $2/3$, then so does the curve $t \mapsto S_t(1, \delta, \delta, 1)$ for small $\delta$. It can be checked that at time $1/3$, the curve $t \mapsto |S_t(1, \delta, \delta, 1)|$ is not $C^2$ differentiable and therefore cannot be an E-spline. □

This counterexample, however, is somewhat degenerate because the transport spline passes through a degenerate measure, and thus it is not clear if the E-spline exists, and if so whether
it remains non-degenerate. We now give another example where the transport spline does not coincide with the E-spline, but the transport spline remains non-degenerate; hence, we believe that the E-spline problem is well-posed for these data.

For this example, we take $\delta > 0$ and let

\[
\mu_0^* = \mu_1^* = \text{uniform on } [-1 + \delta, -1] \cup [1, 1 + \delta], \quad \mu_{1/4}^* = \mu_{3/4}^* = \text{uniform on } [-\delta, \delta].
\]

(B.1)

As in the proof of Proposition 9, $\mathcal{P}_2(\mathbb{R})$ is seen as a convex subset of $L^2[0, 1]$ where probability measures are identified as their quantile function. So our E-spline interpolation can be reformulated as the problem

\[
\inf_{(\mu_t)} \int_0^1 \int_0^1 \| \hat{F}_t^*(u) \|^2 \, du \, dt \quad \text{s.t.} \quad \mu_t = \mu_t^* \text{ for all } t \in \{0, 1/4, 3/4, 1\},
\]

where $F_t^*$ denotes the quantile function of $\mu_t$. In particular, the E-spline interpolation problem can be seen as the transport spline interpolation with the extra constraint that the trajectories of the particles must stay ordered (see Theorem 1).

Denote by $(X_t)$ the random process given by the transport spline problem. One can check that

\[
X_t = \text{sign}(X_0) \left[ \frac{16}{3} (t - 1/2)^2 - \frac{1}{3} + |X_0| - 1 \right].
\]

Clearly, for $\delta$ small enough the quantiles $F_t^*(u)$ of order $u > 1/2$ associated to the transport spline interpolation decrease before $t = 1/4$ and increase after $= 3/4$. In particular, for each $u > 1/2$,
there exists $1/4 < t_u^- < t_u^+ < 3/4$ such that $\partial_t F^i_t(u)|_{t=t_u^-} = \partial_t F^i_t(u)|_{t=t_u^+} = 0$ and $|\partial^2 F^i_t(u)| > 0$ for $t \in (t_u^-, t_u^+)$. One can check then that the function $u \mapsto F^i_t$ at time $t \in [0, 1]$ defined by

$$F^i_t(u) = \begin{cases} F^i_{t_u^+}(u), & u \in (t_u^-, t_u^+) \\ F^i_t(u), & \text{otherwise} \end{cases}$$

is a quantile function. In particular, the measures with quantiles $F^i_t$ interpolate the measures (B.1) and

$$|\partial^2 F^i_t(u)| = \begin{cases} 0, & u \in (t_u^-, t_u^+) \\ |\partial^2 F^i_t(u)|, & \text{otherwise}, \end{cases}$$

ensuring that $F^i_t$ has a lower cost than the transport spline. Thus, the transport spline is not the E-spline. □

Since the transport spline is non-degenerate for this example, we believe that the E-spline also exists and is non-degenerate. Therefore, we expect that the failure of transport splines to equal E-splines in general is not simply due to the fact that E-splines can be ill-posed.

To summarize: when the trajectories of the transport spline remain ordered throughout the interpolation, then it coincides with the E-spline. Otherwise, there is no reason to expect the two notions of spline to coincide.

**APPENDIX C: PROOF OF THE APPROXIMATION GUARANTEE**

Throughout, we assume all random variables are defined on a probability space with probability measure $\mathbb{P}$. Thus, if $X$ is a random variable taking values in $\mathbb{R}^d$, then $\|X\|_{L^2(\mathbb{P})} := \sqrt{\mathbb{E}[\|X\|^2]}$.

We begin by describing the general strategy for proving the approximation guarantee. Consider the interval $[t_{i-1}, t_i]$, let $(X^i_t)$ denote the Lagrangian coupling for $(\mu^i_t)$, and let $(X^*_t)$ be the stochastic process associated with the transport spline. Since $\mu^i_{t_{i-1}} = \mu^i_{t_i}$, we can couple the two processes together so that $X^i_{t_{i-1}} = X^*_{t_{i-1}}$. By the definition of the Wasserstein distance, we can bound $W_2(\mu^i_t, \mu^i_t^*) \leq \|X^i_t - X^*_t\|_{L^2(\mathbb{P})}$, so it suffices to show that the trajectories $(X^i_t)$ and $(X^*_t)$ are close on the interval $[t_{i-1}, t_i]$.

We will use a basic deterministic fact: if two curves $x$ and $y$ defined on $[0, \delta]$ are such that:

- $x(0) = y(0)$,
- $\dot{x}(0) = \dot{y}(0) + O(\delta)$, and
- the two curves satisfy the curvature bound
  $$\sup_{t \in [0, \delta]} \{ \|\dot{x}(t)\| \vee \|\dot{y}(t)\| \} \leq R,$$

then it follows that $\sup_{t \in [0, \delta]} \|x(t) - y(t)\| \leq CR\delta^2$, where $C$ is a numerical constant.

1. the velocities of $X^i_t$ and $X^*_t$ at time $t = t_{i-1}$ are within $O(\delta)$ of each other (Proposition 13);
2. the trajectory $(X^i_t)$ has curvature $O(R)$ (Proposition 14);
3. the trajectory $(X^*_t)$ has curvature $O(R)$;

The last step is immediate from our assumptions; the point of the second step is to control the curvature of the interpolated process $(X^i_t)$ in terms of the curvature of the true process $(X^*_t)$.

Putting these pieces together, we give the proof of Theorem 2 in Appendix C.4.
C.1 Notation

Since we study the approximation guarantee in the Bures-Wasserstein setting, we can equivalently think in terms of the probability measure (a Gaussian), or in terms of the covariance matrix. It will be useful to employ the language of matrices, so we fix notational conventions here.

Associated with the curve \((\mu^*_t)\), we have a corresponding curve of covariance matrices \((\Sigma_t)\) such that \(\mu^*_t = \mathcal{N}(0, \Sigma_t)\).

Given a matrix \(A \in \mathbb{R}^{d \times d}\), we define the norm

\[ \|A\|_\Sigma := \sqrt{\langle A, \Sigma A \rangle}. \]

The norm is defined so that if \(X^* \sim \mathcal{N}(0, \Sigma)\), then \(\|AX^*\|_{L^2(\mathbb{P})} = \|A\|_\Sigma\). From our eigenvalue bound we have \(\|A\|_\Sigma \geq \sqrt{\lambda_{\min}(\Sigma)} \|A\|_F\).

The Monge map \(T\) between two Gaussians is the linear map \(T(X)\) given in (5.1) and abusing notation slightly, we identify the map \(T\) with the corresponding matrix, and we write \(T(x) = Tx\).

In particular, linearity of the Monge maps implies that the velocity vector field \((v^*_t)\) associated to the Lagrangian coupling of the curve, is also linear for each \(t\): \(v^*_t\) is a symmetric linear mapping \(\mathbb{R}^d \rightarrow \mathbb{R}^d\), that is, there exists a symmetric matrix \(V^*_t \in \mathbb{R}^{d \times d}\) such that \(v^*_t(x) = V^*_tx\).

C.2 Control of the Velocities

We write \(\delta_i := t_{i+1} - t_i\) and \(\delta := \max_{i \in [N]} \delta_i\). The first step is to prove a quantitative bound on how well the Monge map \(T_i\) approximates \(\text{id} + \delta_i v_{t_{i-1}}\). We prove a more general approximation result which may be of independent interest.

**Theorem 3.** Let \(t, t + h \in [0, 1]\), where \(h \neq 0\). Write \(\delta := |h|\) and assume \(\delta \leq c\sqrt{\lambda_{\min}(\Sigma_t)}/L\), for some constant \(0 < c < 1\). Let \(T\) denote the Monge map from \(\mu^*_t\) to \(\mu^*_{t+h}\), and let \(\bar{T} : \mathbb{R}^d \rightarrow \mathbb{R}^d\) be another linear mapping satisfying the following properties:

1. \(\bar{T}\) can be identified with a symmetric matrix.
2. \(\|TX_t^* - X_t^*\|_{L^2(\mathbb{P})} \leq c\sqrt{\lambda_{\min}(\Sigma_t)}\).

Then,

\[ \|TX_t^* - \bar{T}X_t^*\|_{L^2(\mathbb{P})} \leq \frac{1 + 2c}{1 - c} \|TX_t^* - X_{t+h}^*\|_{L^2(\mathbb{P})}. \]

**Proof.** Let \(e := X_{t+h}^* - \bar{T}X_t^*\).

Consider the quadratic function \(\varphi : \mathbb{R}^d \rightarrow \mathbb{R}\) defined by \(\varphi(x) := \langle x, Ax \rangle\), where \(A := (T - \bar{T})/\|T - \bar{T}\|_{\Sigma_t}\). Note that \(A\) is symmetric (since \(T\) and \(\bar{T}\) are). Then,

\[ \mathbb{E}\varphi(TX_t^*) = \mathbb{E}\varphi(X_{t+h}^*) = \mathbb{E}\varphi(\bar{T}X_t^* + e). \]

Expanding this out,

\[ 0 = \mathbb{E}\langle (T + \bar{T})X_t^* + e, A\{(T - \bar{T})X_t^* - e\}\rangle \]
\[ = \mathbb{E}\langle (T + \bar{T})X_t^*, A(T - \bar{T})X_t^* \rangle + \text{error}. \]
We next bound the error term. First, note that by our assumption,

\[ ||T - I_d||_{\Sigma_t} = W_2(\mu_t^*, \mu_{t+h}^*) \leq L\delta \leq c\sqrt{\lambda_{\text{min}}}, \]

\[ ||\bar{T} - I_d||_{\Sigma_t} \leq c\sqrt{\lambda_{\text{min}}}, \]

where we write \( \lambda_{\text{min}} = \lambda_{\text{min}}(\Sigma_t) \). The error term is split into two further terms. For the first term,

\[ |E(e, A(T - \bar{T})X_t^*)| \leq ||e||_{L^2(P)} ||A(T - \bar{T})||_{\Sigma_t} \]

\[ \leq ||e||_{L^2(P)} ||A||_F ||T - \bar{T}||_{\Sigma_t} \]

\[ \leq ||e||_{L^2(P)} \frac{1}{\sqrt{\lambda_{\text{min}}}} (||T - I_d||_{\Sigma_t} + ||\bar{T} - I_d||_{\Sigma_t}) \]

\[ \leq 2c ||e||_{L^2(P)}, \]

where we used the fact that \( ||A||_{\Sigma_t} \leq 1 \) implies that \( ||A||_F \leq 1/\sqrt{\lambda_{\text{min}}} \). The second term is bounded by

\[ |E(\langle(T + \bar{T})X_t^* + e, Ae\rangle) - |E(\langle T X_t^* + X_{t+h}^* - 2X_t^*, Ae \rangle) + 2|E(\langle X_t^*, Ae \rangle)\rangle| \]

\[ \leq \{ ||A||_F (||T - I_d||_{\Sigma_t} + ||X_{t+h}^* - X_t^*||_{L^2(P)} + 2||A||_{\Sigma_t}) ||e||_{L^2(P)} \}

\[ \leq 2(1 + c) ||e||_{L^2(P)}, \]

where we used

\[ ||X_{t+h}^* - X_t^*||_{L^2(P)}^2 = E\left[ \left| \int_t^{t+h} \dot{X}_s^* \, ds \right|^2 \right] \leq \delta \int_t^{t+h} ||\dot{X}_s^*||_{L^2(P)}^2 \, ds \leq L^2\delta^2. \]

Thus, we have

\[ 2||T - \bar{T}||_{\Sigma_t} = 2E(\langle X_t^*, A(T - \bar{T})X_t^* \rangle) \]

\[ = -E(\langle(T + \bar{T} - 2I_d)X_t^*, (T - \bar{T})X_t^* \rangle) + \text{error} \]

\[ \leq (||T - I_d||_{\Sigma_t} + ||\bar{T} - I_d||_{\Sigma_t}) ||A||_F ||T - \bar{T}||_{\Sigma_t} + \text{error} \]

\[ \leq 2c ||T - \bar{T}||_{\Sigma_t} + (2 + 4c) ||e||_{L^2(P)} \]

which finally yields

\[ ||T - \bar{T}||_{\Sigma_t} \leq \frac{1 + 2c}{1 - c} ||e||_{L^2(P)} \]

as required. \( \square \)

**Corollary 1.** Let \( t, t + h \in [0, 1] \), where \( h \neq 0 \), and write \( \delta := |h| \). Let \( k \in \{0, 1, 2\} \), and suppose \( \delta \) is small enough so that

\[ \sum_{i=1}^{k} \frac{R_i\delta^i}{i!} \leq c\sqrt{\lambda_{\text{min}}(\Sigma_t)}, \]

where we set \( R_i := \sup_{t \in [0, 1]} ||\partial^i X_t^*||_{L^2(P)} \). Then,

\[ \left| \left| TX_t^* - \sum_{i=0}^{k} \frac{h^i}{i!} (\partial^i X_t^*) \right| \right|_{L^2(P)} \leq \frac{1 + 2c}{1 - c} \frac{R_{k+1}\delta^{k+1}}{(k + 1)!}. \]
Proof. We apply Theorem 3 with

$$\bar{T}X^*_t = \sum_{i=0}^{k} \frac{h^i}{i!} (\partial^i X^*)_t.$$ 

Using $\dot{X}^*_t = V^*_t X^*_t$, where $V^*_t$ is symmetric, we obtain:

$$\dot{X}^*_t = V^*_t X^*_t, $$
$$\ddot{X}^*_t = \dot{V}^*_t X^*_t + V^*_t X^*_t = (\dot{V}^*_t + V^*_t^2)X^*_t, $$
$$\dddot{X}^*_t = (\ddot{V}^*_t + 2\dot{V}^*_t V^*_t + V^*_t^3)X^*_t, $$
$$\vdots$$

Observe that the $i$th derivative of $t \mapsto X^*_t$ at $t$ is indeed a linear function of $X^*_t$, but for $i \geq 3$ it is no longer given by a symmetric matrix, so it no longer satisfies the first assumption of Theorem 2; this is why we restrict ourselves to $k = 0, 1, 2$.

For the third assumption of Theorem 2, note that

$$\|\bar{T}X^*_t - X^*_t\|_{L^2(P)} = \left\| \sum_{i=1}^{k} \frac{h^i}{i!} (\partial^i X^*)_t \right\|_{L^2(P)} \leq \sum_{i=1}^{k} \frac{\delta^i R_i}{i!} \leq c\sqrt{\lambda_{\min}(\Sigma_t)},$$

by our assumption on $\delta$.

Finally, the error $e := X^*_{t+h} - \bar{T}X^*_t$ is controlled via Taylor’s theorem:

$$\|e\|_{L^2(P)} = \left\| X^*_{t+h} - \sum_{i=0}^{k} \frac{h^i}{i!} (\partial^i X^*)_t \right\|_{L^2(P)}$$
$$= \left\| \int_t^{t+h} (\partial^{k+1} X^*)_s \frac{(s-t)^k}{k!} ds \right\|_{L^2(P)}$$
$$\leq \frac{R_{k+1} \delta^{k+1}}{(k+1)!}.$$ \hfill \Box

Remark 3. If we let $\delta \searrow 0$, we can also take $c \searrow 0$, obtaining

$$\limsup_{\delta \searrow 0} \frac{1}{\delta^{k+1}} \left\| T X^*_t - \sum_{i=0}^{k} \frac{h^i}{i!} (\partial^i X^*)_t \right\|_{L^2(P)} \leq \frac{R_{k+1}}{(k+1)!}.$$ 

Comparing this to a Euclidean Taylor expansion, this is apparently sharp.

Corollary 1 says that in order to prove our desired result $\dot{X}_{t-1} = \dot{X}^*_{t-1} + O(\delta)$, it suffices to show that $\dot{X}_{t-1} = (T_1 X_{t-1} - X_{t-1})/\delta_t + O(\delta)$ (since the RHS of both expressions equals $V^*_{t-1} X_{t-1} = V^*_{t-1} X^*_{t-1}$ up to $O(\delta)$). Since the latter statement involves only the process $(X_t)$, it is easier to prove.
However, there is still a major difficulty to overcome: $\dot{X}_{t_{i-1}}$ is the velocity of an interpolating cubic spline, which depends on all of the interpolated points $X_{t_0}, X_{t_1}, \ldots, X_{t_N}$. In Appendix D, we show that the derivative of the cubic spline interpolation can be understood in terms of the linear system of equations involving the quantities

$$
\Delta_i := \frac{X_{t_{i+1}} - X_{t_i}}{\delta_{i+1}} - \frac{X_{t_i} - X_{t_{i-1}}}{\delta_i}, \quad i \in [N - 1].
$$

Therefore, we next control these quantities.

**Proposition 12.** Assume $\delta \leq \frac{\sqrt{\lambda_{\min}}}{2L}$. For each $i \in [N - 1]$, it holds that

$$
\|\frac{X_{t_{i+1}} - X_{t_i}}{\delta_{i+1}} - \frac{X_{t_i} - X_{t_{i-1}}}{\delta_i}\|_{L^2(\mathcal{P})} \leq \frac{25}{4} R\delta.
$$

**Proof.** From Corollary 1,

$$
\|\frac{X_{t_i} - X_{t_{i-1}}}{\delta_i} - V_{t_{i-1}}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})} = \|\frac{T_i - I_d}{\delta_i} - V_{t_{i-1}}^*\|_{\Sigma_{t_{i-1}}} \leq 2R\delta_i,
$$

where we use the fact that $X_{t_{i-1}} \sim \mu_{t_{i-1}}^*$ and that $X_{t_i} = T_i X_{t_{i-1}}$. Similarly,

$$
\|\frac{X_{t_{i+1}} - X_{t_i}}{\delta_{i+1}} - V_{t_i}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})} \leq 2R\delta_{i+1}.
$$

Therefore,

$$
\|\Delta_i\|_{L^2(\mathcal{P})} \leq 4R\delta + \|V_{t_i}^* X_{t_{i-1}} - V_{t_{i-1}}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})}.
$$

Since $X_{t_i} = T_i X_{t_{i-1}}$, we replace $T_i$ by $I_d + \delta_i V_{t_{i-1}}^*$.

$$
\|V_{t_i}^* X_{t_{i-1}} - V_{t_{i-1}}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})} \\
\leq \|V_{t_i}^* (T_i - I_d - \delta_i V_{t_{i-1}}^*) X_{t_{i-1}}\|_{L^2(\mathcal{P})} + \|V_{t_i}^* (I_d + \delta_i V_{t_{i-1}}^*) X_{t_{i-1}} - V_{t_{i-1}}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})}.
$$

We control the first term using Corollary 1:

$$
\|V_{t_i}^* (T_i - I_d - \delta_i V_{t_{i-1}}^*) X_{t_{i-1}}\|_{L^2(\mathcal{P})} \leq \|V_{t_i}^*\|_F \|((T_i - I_d - \delta_i V_{t_{i-1}}^*) X_{t_{i-1}}\|_{L^2(\mathcal{P})} \\
\leq \frac{L}{\sqrt{\lambda_{\min}}} \|T_i - I_d - \delta_i V_{t_{i-1}}^*\|_{\Sigma_{t_{i-1}}} \\
\leq \frac{L}{\sqrt{\lambda_{\min}}} \cdot 2R\delta_i^2 \leq R\delta_i,
$$

where we used $\|V_{t_i}^*\|_F \leq \lambda_{\min}^{-1/2} \|V_{t_i}^*\|_{\Sigma_{t_{i-1}}} \leq L\lambda_{\min}^{-1/2}$ by our Lipschitz assumption. Now for the second term. Introduce the random trajectory $(X_t^*)$ sampled from the true curve $(\mu_t^*)$ with the Lagrangian coupling, and couple the process $(X_t)$ with $(X_t^*)$ by setting $X_{t_{i-1}} = X_{t_{i-1}}^*$. Thus,

$$
\|V_{t_i}^* (I_d + \delta_i V_{t_{i-1}}^*) X_{t_{i-1}} - V_{t_{i-1}}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})} \\
\leq \|V_{t_i}^* X_{t_{i-1}} - V_{t_{i-1}}^* X_{t_{i-1}}\|_{L^2(\mathcal{P})} + \|V_{t_i}^* ((I_d + \delta_i V_{t_{i-1}}^*) X_{t_{i-1}} - X_{t_{i-1}}^*)\|_{L^2(\mathcal{P})}.
$$
It is easy to control
\[
\|V_{t_i}^* X_{t_i}^* - V_{t_i-1}^* X_{t_i-1}^* \|_{L^2(P)} = \left\| \int_{t_{i-1}}^{t_i} \ddot{X}_t^* \, dt \right\|_{L^2(P)} \leq R \delta_i.
\]
Lastly,
\[
\|V_{t_i}^* \{(I_a + \delta_i V_{t_i-1}^*) X_{t_i-1}^* - X_{t_i-1}^* \} \|_{L^2(P)} \leq \|V_{t_i}^*\|_{F} \| X_{t_i}^* - X_{t_i-1}^* - \delta_i V_{t_i-1}^* X_{t_i-1}^* \|_{L^2(P)}
\leq \frac{L}{\sqrt{\lambda_{\min}}} \left( \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} \ddot{X}_s^* ds \, dt \right)_{L^2(P)}
\leq \frac{L}{\sqrt{\lambda_{\min}}} \cdot \frac{R \delta_i^2}{2} \leq \frac{R \delta_i}{4}.
\]
Putting it all together, we obtain
\[
\|\Delta_i\|_{L^2(P)} \leq \frac{25}{4} R \delta. \quad \square
\]

To match notation with Appendix D, we set
\[
M_i := \ddot{X}_{t_i-1}, \quad i \in [N+1].
\]

**Lemma 2.** Assume \( \delta \leq \sqrt{\lambda_{\min}}/(2L) \). It holds that
\[
\|M_i\|_{L^2(P)} \leq \frac{75(1 + \alpha)^2}{4 \alpha^3} R.
\]

**Proof.** As described in Appendix D, we know that \( M = 6T^{-1} \Delta \), where the entries of \( T^{-1} \) are bounded in Lemma 3. Thus,
\[
\|M_i\|_{L^2(P)} = 6 \left\| \sum_{j=1}^{N-1} (T^{-1})_{i,j} \Delta_j \right\|_{L^2(P)}
\leq 6 \sum_{j=1}^{N-1} \| (T^{-1})_{i,j} \| \| \Delta_j \|_{L^2(P)}
\leq 6 \sum_{j=1}^{N-1} \frac{1}{4 \alpha^2 \delta} \frac{1}{(1 + \alpha)^{|i-j|-1}} \frac{25}{4} R \delta
\leq \frac{75R}{4 \alpha^2 \sum_{k=0}^{\infty} (1 + \alpha)^{k-1}} = \frac{75(1 + \alpha)^2}{4 \alpha^3} R,
\]
where we use Proposition 12. \( \square \)

Finally, we are ready to state our control on the velocity of the trajectory \((X_t)\).
Proposition 13. Assume \( \delta \leq \sqrt{\lambda_{\text{min}}/(2L)} \). Then,
\[
\|X_{t_{i-1}} - \dot{X}_{t_{i-1}}^*\|_{L^2(P)} \leq \frac{16\alpha^3 + 75(1 + \alpha)^2}{8\alpha^3} R\delta.
\]

Proof. It holds that
\[
\dot{X}_{t_{i-1}} = \frac{X_{t_i} - X_{t_{i-1}}}{\delta_i} - \frac{M_{i+1} + 2M_i}{6} \delta_i
\]
(see Appendix D). Therefore,
\[
\|\dot{X}_{t_{i-1}} - \frac{X_{t_i} - X_{t_{i-1}}}{\delta_i}\|_{L^2(P)} \leq \frac{\|M_{i+1}\|_{L^2(P)} + 2\|M_i\|_{L^2(P)}}{6} \delta \leq \frac{75(1 + \alpha)^2}{8\alpha^3} R\delta,
\]
by Lemma 2. Next, we recall that \( X_{t_i} = T_iX_{t_{i-1}} \), and that \( (X_t) \) and \( (X_t^*) \) are coupled so that \( X_{t_{i-1}} = X_{t_{i-1}}^* \). Thus,
\[
\|\dot{X}_{t_{i-1}} - \dot{X}_{t_{i-1}}^*\|_{L^2(P)} \leq \|\dot{X}_{t_{i-1}} - \frac{T_iX_{t_{i-1}} - X_{t_{i-1}}}{\delta_i}\|_{L^2(P)} + \|\dot{X}_{t_{i-1}}^* - \frac{T_iX_{t_{i-1}}^* - X_{t_{i-1}}^*}{\delta_i}\|_{L^2(P)}
\]
\[
\leq \frac{75(1 + \alpha)^2}{8\alpha^3} R\delta + 2R\delta,
\]
where we invoke Corollary 1 again. \( \square \)

C.3 Curvature of the Transport Spline

Next, we must bound the curvature of \( (X_t) \), but this is an easy task given what we have established so far.

Proposition 14. Assume \( \delta \leq \sqrt{\lambda_{\text{min}}/(2L)} \). Then,
\[
\sup_{t \in [0,1]} \|\ddot{X}_t\|_{L^2(P)} \leq \frac{75(1 + \alpha)^2}{4\alpha^3} R.
\]

Proof. Indeed, \( t \mapsto \ddot{X}_t \) is a piecewise linear function (see Appendix D), so it is maximized at the knots. For \( t \in [t_{i-1}, t_i] \), it follows that
\[
\|\ddot{X}_t\|_{L^2(P)} = \left\| \frac{t_i - t}{\delta_i} \ddot{X}_{t_{i-1}} + \frac{t - t_{i-1}}{\delta_i} \ddot{X}_{t_i} \right\|_{L^2(P)}
\]
\[
\leq \frac{t_i - t}{\delta_i} \|\ddot{X}_{t_{i-1}}\|_{L^2(P)} + \frac{t - t_{i-1}}{\delta_i} \|\ddot{X}_{t_i}\|_{L^2(P)}
\]
\[
\leq \|\ddot{X}_{t_{i-1}}\|_{L^2(P)} \lor \|\ddot{X}_{t_i}\|_{L^2(P)}
\]
\[
= \|M_i\|_{L^2(P)} \lor \|M_{i+1}\|_{L^2(P)}
\]
\[
\leq \frac{75(1 + \alpha)^2}{4\alpha^3} R,
\]
by Lemma 2. \( \square \)
C.4 Proof of the Main Theorem

**Proof of Theorem 2.** Let \( t \in [t_{i-1}, t_i] \), and let the processes \((X_t)\) and \((X_t^*)\) be coupled with \(X_{t_{i-1}} = X_{t_{i-1}}^*\). Then,

\[
\|X_t - X_t^*\|_{L^2(\mathbb{P})} \leq \delta_i \|\dot{X}_{t_{i-1}} - \dot{X}_{t_{i-1}}^*\|_{L^2(\mathbb{P})} + \left\| \int_{t_{i-1}}^{t} \int_{t_{i-1}}^{t} (\dddot{X}_s - \dddot{X}_s^*) \, ds \, dt \right\|_{L^2(\mathbb{P})}
\]

\[
\leq \frac{16\alpha^3 + 75(1 + \alpha)^2}{8\alpha^3} R\delta^2 + \frac{\delta^2}{2} \sup_{t \in [0, 1]} (\|\dddot{X}_t\|_{L^2(\mathbb{P})} + \|\dddot{X}_t^*\|_{L^2(\mathbb{P})})
\]

\[
\leq \frac{10\alpha^3 + 75(1 + \alpha)^2}{4\alpha^3} R\delta^2 \leq \frac{115}{2\alpha^3} R\delta^2,
\]

where we have used Proposition 13 and Proposition 14. □

C.5 Piecewise Geodesic Interpolation

In this section, we study the approximation error of piecewise geodesic interpolation. Namely, we define a stochastic process, still denoted \((X_t)\), as follows.

1. Draw \(X_{t_0} \sim \mu_{t_0}\).
2. For \(i = 1, \ldots, N\), set \(X_{t_i} := T_i(X_{t_{i-1}})\).
3. We join the points \(X_{t_0}, X_{t_1}, \ldots, X_{t_N}\) via straight lines. Namely, for \(t \in [t_{i-1}, t_i]\) we set

\[
X_t = \frac{t - t_i}{t_i - t_{i-1}} X_{t_{i-1}} + \frac{t_i - t}{t_i - t_{i-1}} X_{t_i}.
\]

Let \(\mu_t\) denote the law of \(X_t\).

**Theorem 4.** Let the notation and assumptions of Theorem 2 hold (except for the definition of \((\mu_t)\)). Then,

\[
\sup_{t \in [0, 1]} W_2(\mu_t, \mu_t^*) \leq \frac{5}{2} R\delta^2.
\]

**Proof.** As in Appendix C.4, we have

\[
\|X_t - X_t^*\|_{L^2(\mathbb{P})} \leq \delta_i \|\dot{X}_{t_{i-1}} - \dot{X}_{t_{i-1}}^*\|_{L^2(\mathbb{P})} + \left\| \int_{t_{i-1}}^{t} \int_{t_{i-1}}^{t} \dddot{X}_s^* \, ds \, dt \right\|_{L^2(\mathbb{P})}
\]

\[
\leq 2R\delta^2 + \frac{1}{2} R\delta^2.
\]

Here, we use several facts: (1) \(\dot{X}_{t_{i-1}}^*\), the derivative of \((X_t)_t\) at \(t_{i-1}\) from the right, equals

\[(T_i - I_d)X_{t_{i-1}} = (T_i - I_d)X_{t_{i-1}}^*;
\]

and so we can apply Corollary 1; (2) the curve \((X_t)\), consisting of piecewise straight lines, has no acceleration. This finishes the proof. □
Formally, Theorem 4 is a slightly better approximation guarantee than Theorem 2. Theorem 4 can also be strengthened asymptotically to

$$\limsup_{\delta \to 0} \frac{1}{\delta^2} \sup_{t \in [0,1]} W_2(\mu_{\delta,t}, \mu_{\star,t}) \leq R,$$

as in Section 5.2. Of course, we do not advocate for using piecewise geodesic interpolation because it is unsuitable for trajectory estimation (see Figure 1).

**APPENDIX D: NATURAL CUBIC SPLINES**

For the reader’s convenience and to make the paper more self-contained, in this section we present a derivation of natural cubic splines and some of their properties. The results obtained here are used in Appendix C for the proof of the main approximation result (Theorem 2).

We are given times $0 = t_0 < t_1 < \cdots < t_N = 1$ and corresponding points $(x_{t_0}, x_{t_1}, \ldots, x_{t_N})$ in $\mathbb{R}^d$. Our goal is to construct a piecewise cubic polynomial interpolation $y : [0,1] \to \mathbb{R}^d$ which is $C^2$ smooth.

We parametrize $y$ in the following way: for each $i \in [N]$ and for $t \in [t_{i-1}, t_i]$, we set $y(t) = y_i(t)$, where

$$y_i(t) = a_i (t - t_{i-1})^3 + b_i (t - t_{i-1})^2 + c_i (t - t_{i-1}) + d_i.$$

Computing derivatives,

$$x_{t_{i-1}} = y_i(t_{i-1}) = d_i,$$

$$x_{t_i} = y_i(t_i) = a_i \delta_i^3 + \frac{m_i}{2} \delta_i^2 + c_i \delta_i + d_i,$$

$$\dot{y}_i(t_{i-1}) = c_i,$$

$$\dot{y}_i(t_i) = 3a_i \delta_i^2 + m_i \delta_i + c_i,$$

$$\ddot{y}_i(t_{i-1}) = m_i,$$

$$\ddot{y}_i(t_i) = 6a_i \delta_i + m_i,$$

where define $\delta_i := t_i - t_{i-1}$ and $m_i := 2b_i$ (and anticipating the natural boundary condition, which asserts $\ddot{y}(0) = \ddot{y}(1) = 0$, we make the convention $m_{N+1} := 0$). Using continuity of the first and second derivatives of $y$ at the knots, we solve for the coefficients of the polynomial $y_i$ in terms of the variables $m$ and $x$:

$$a_i = \frac{m_{i+1} - m_i}{6 \delta_i},$$

$$b_i = \frac{m_i}{2},$$

$$c_i = \frac{x_{t_i} - x_{t_{i-1}}}{\delta_i} - \frac{m_{i+1} + 2m_i}{6} \delta_i,$$

$$d_i = x_{t_{i-1}}.$$

Therefore, it suffices to work with the variables $m$. 
If we plug these equations back into the continuity condition for the first derivative at the knot, after some algebra we obtain the equations

\[ 6\Delta_i = \delta_i m_i + 2(\delta_i + \delta_{i+1})m_{i+1} + \delta_{i+2}m_{i+2}, \quad i = 1, \ldots, N - 1, \]

where we have defined the quantities

\[ \Delta_i := \frac{x_{t_{i+1}} - x_{t_i}}{\delta_i} - \frac{x_{t_i} - x_{t_{i-1}}}{\delta_i}, \]

a proxy for the second derivative of the data points.

We can express these equations in matrix form (including also the natural boundary condition \( m_1 = 0 \)):

\[
\begin{bmatrix}
2(\delta_1 + \delta_2) & \delta_2 & & & \\
\delta_2 & \ddots & \ddots & & \\
& \ddots & \ddots & \delta_{N-1} & \\
& & \delta_{N-1} & 2(\delta_{N-1} + \delta_N) \\
\end{bmatrix}
\begin{bmatrix}
m_2 \\
m_3 \\
\vdots \\
m_N \\
\end{bmatrix}
= 6\Delta.
\]

The matrix \( T \) above is a symmetric tridiagonal matrix of size \( N - 1 \).\(^9\) To obtain bounds on \( m \), we will study the inverse \( T^{-1} \) of \( T \).

**Lemma 3.** Assume that for each \( i \in [N] \), we have \( \alpha \delta \leq t_i - t_{i-1} \leq \delta \). Then, we have the entrywise bound

\[ |(T^{-1})_{i,j}| \leq \frac{1}{4\alpha^2 (1 + \alpha)^{|i-j|-1}} \frac{1}{\delta}, \quad i, j \in [N - 1]. \]

**Proof.** We write \( T = B + D \), where

\[
B := \begin{bmatrix}
0 & \delta_2 & & & \\
\delta_2 & \ddots & \ddots & & \\
& \ddots & \ddots & \delta_{N-1} & \\
& & \delta_{N-1} & 0 \\
\end{bmatrix},
\]

\[
D := 2\text{diag}(\delta_1 + \delta_2, \ldots, \delta_{N-1} + \delta_N).
\]

Therefore,

\[
T^{-1} = (B + D)^{-1}
= D^{-1/2}(I_{N-1} + D^{-1/2}BD^{-1/2})^{-1}D^{-1/2}
= \sum_{k=0}^{\infty} (-1)^k D^{-1/2} \underbrace{(D^{-1/2}BD^{-1/2})^k}_{= M} D^{-1/2}.
\]

\(^9\)To be precise, we should write this as the block matrix equation \((T \otimes I_d)m = 6\Delta\).
The matrix $M$ is

$$
M = \begin{bmatrix}
0 & \gamma_2 & & \\
\gamma_2 & \ddots & \ddots & \\
& \ddots & \ddots & \gamma_{N-1} \\
& & \gamma_{N-1} & 0
\end{bmatrix},
$$

where we set

$$
\gamma_i := \frac{\delta_i}{2\sqrt{(\delta_{i-1} + \delta_i)(\delta_i + \delta_{i+1})}} \leq \frac{1}{2(1 + \alpha)}.
$$

Since $M$ has non-negative entries, we have the entrywise bound

$$
M^k \leq \frac{1}{\{2(1 + \alpha)\}^k} \begin{bmatrix}
0 & 1 & & \\
1 & \ddots & \ddots & \\
& \ddots & \ddots & 1 \\
& & 1 & 0
\end{bmatrix}^k \quad := A.
$$

The matrix $A$ is the adjacency matrix of the path graph on $\{1, \ldots, N-1\}$, so $(A^k)_{i,j}$ is the number of paths from $i$ to $j$ of length $k$. We can trivially bound this number by $2^k \mathbbm{1}_{|i-j| \leq k}$. From this we deduce the entrywise bound

$$
(M^k)_{i,j} \leq \frac{1}{(1 + \alpha)^k} \mathbbm{1}_{|i-j| \leq k}.
$$

Therefore,

$$
|(T^{-1})_{i,j}| \leq \sum_{k=0}^{\infty} \frac{1}{2\sqrt{(\delta_i + \delta_{i+1})(\delta_j + \delta_{j+1})}} \frac{1}{(1 + \alpha)^k} \mathbbm{1}_{|i-j| \leq k} \\
\leq \frac{1}{4\alpha\delta} \sum_{k=|i-j|}^{\infty} \frac{1}{(1 + \alpha)^k} = \frac{1}{4\alpha^2 (1 + \alpha)^{|i-j|-1}} \frac{1}{\delta}. \quad \Box
$$

**APPENDIX E: DETAILS FOR THE EXPERIMENTS**

In this section we provide further details for the experiments in the paper, except for the thin-plate spline example (which is discussed in Appendix F).

**E.1 Figure 1**

In this figure, we set five Gaussians as our interpolation knots, alternating between

$$
\mathcal{N}\left(\begin{bmatrix}7(k-1) \\ 0\end{bmatrix}, \begin{bmatrix}4 & 0 \\ 0 & 2\end{bmatrix}\right) \quad \text{for } k \text{ odd}
$$
and
\[ \mathcal{N}\left(\begin{bmatrix} 7(k-1) \\ 7 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right) \]
for \( k \) even,

where \( k = 1, \ldots, 5 \).

To determine the linear and cubic spline interpolations we first computed the optimal transport maps between the neighboring Gaussians. The closed-form formula for the Monge map from \( \mathcal{N}(\mu_1, \Sigma_1) \) to \( \mathcal{N}(\mu_2, \Sigma_2) \) is
\[
T(x) = \mu_2 + A(x - \mu_1), \quad A = \Sigma_1^{-\frac{1}{2}}(\Sigma_2 \Sigma_1^{-\frac{1}{2}})^{\frac{1}{2}} \Sigma_1^{-\frac{1}{2}}.
\]

The gray lines in both figures show the trajectories of individual sample points along our interpolations. To draw them, we obtained a sample \( X_0 \) from the Gaussian at time \( t = 0 \), repeatedly applied the Monge maps between successive Gaussians in time, and fit a piecewise linear or natural cubic spline through these points as described in Section 4.1.

Since the maps between successive Gaussians are linear and the formula for the linear or natural cubic spline is linear in its knots, the value of the spline \( S_t(X_0) \) interpolation at time \( t \) is linear in \( X_0 \). Hence, given the covariance of the Gaussian at time \( t = 0 \), we used this linear map \( S_t \) to compute the covariance of the interpolated Gaussian at time \( t \). Likewise, by taking a linear or cubic spline through the means of the Gaussians at the knot points, we obtained the means of the interpolated Gaussians at any given time. Using this information, we plotted the interpolated Gaussians at the halfway points between the knots for both the linear and cubic spline interpolations.

E.2 Figure 3

To simulate the \( n \)-body trajectories, we used the Python nBody simulator by Cabrera & Li, which can be accessed at https://github.com/GabrielSCabrera/nBody.

We created 15 smaller bodies, each of mass \( 5 \times 10^9 \) and radius 1. Each body was initialized with a position \( x \) and a velocity \( v \) drawn randomly according to
\[
x \sim \mathcal{N}\left(\begin{bmatrix} 100 \\ \end{bmatrix}, \begin{bmatrix} 30 & 0 \\ 0 & 20 \end{bmatrix}\right), \quad v \sim \mathcal{N}\left(\begin{bmatrix} 10 \\ -20 \end{bmatrix}, \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}\right).
\]

In addition, we also created one larger body, with mass \( 10^{11} \) and radius 10, initialized at the origin with no initial velocity.

We simulated the trajectories of the planets for 5 seconds sampled every 0.02 seconds. We took the positions of the bodies at 5 evenly spaced times as the knots for our interpolation. In order to solve the matching problem between planets at neighboring knot times, we placed a uniform empirical distribution over the planets at both times and used the Python Optimal Transport (POT) library function `ot.emd` to compute the Monge map between these two distributions. We checked post process that the Monge maps computed were indeed valid matchings (i.e. permutation matrices).

Given the Monge maps between knots, we applied Algorithm 1 to interpolate the empirical distributions of the bodies using cubic splines. Note that in our cubic spline reconstruction, it is possible to observe mistakes in the matching, i.e., the Monge map may not necessarily map a body at one time to the same body at a future time. Such mismatches seem unavoidable without using a more sophisticated method which takes into account the physical model in the simulation.
APPENDIX F: FURTHER DETAILS FOR THIN-PLATE SPLINES

F.1 Simultaneously Optimal Coupling

In Section 6 we introduce the following coupling. Let $U$ be a uniform random variable on $[0, 1]$, and set

$$X_{x_i} = F_{u_i}^{-1}(U), \quad i = 0, 1, \ldots, N.$$ 

Then, $(X_{x_0}, X_{x_1}, \ldots, X_{x_N})$ is a simultaneously optimal coupling of the measures $\mu_{x_0}^*, \mu_{x_1}^*, \ldots, \mu_{x_N}^*$. This follows directly from Santambrogio (2015, §2.1-2.2), but we provide some additional explanation here.

As described in Section 2, the Monge map $T_{i,j}$ from $\mu_{x_i}^*$ to $\mu_{x_j}^*$ is characterized as the $(\mu_{x_i}^*-a.e.)$ unique mapping which both pushes $\mu_{x_i}^*$ forward to $\mu_{x_j}^*$ and is the gradient of a convex function. In one dimension, the latter condition simply means that $T_{i,j}$ is an increasing function. It is easily checked that $F_{\mu_{x_j}}^{-1} \circ F_{\mu_{x_i}}$ satisfies these properties, and thus

$$T_{i,j} = F_{\mu_{x_j}}^{-1} \circ F_{\mu_{x_i}}.$$

Now, observe that a composition of increasing maps is increasing, which implies that $T_{j,k} \circ T_{i,j}$ must be the Monge map $T_{i,k}$. This key fact directly implies the existence of the simultaneously optimal coupling of the measures. In higher dimensions, this breaks down because the composition of Monge maps is no longer necessarily a Monge map (that is, the composition of gradients of functions is not necessarily the gradient of a function).

F.2 Gaussian Splines and Quantiles

Recall that the $\alpha$-quantile of a measure $\mu$ is the value $c_\alpha$ for which $\mu((-\infty, c_\alpha]) = \alpha$. If $\mu$ has CDF $F_{\mu}$, then the $\alpha$-quantile is simply $F_{\mu}^{-1}(\alpha)$. If we denote by $\Phi$ the CDF of the standard Gaussian distribution, then the $\alpha$-quantile of $\mathcal{N}(0, 1)$ is $\Phi^{-1}(\alpha)$, and the $\alpha$-quantile of $\mathcal{N}(m, \sigma^2)$ is $m + \Phi^{-1}(\alpha)\sigma$.

Suppose the measures $\mu_{x_i}^*$, $i = 0, 1, \ldots, N$, are all one-dimensional Gaussians, and write $\mu_{x_i}^* = \mathcal{N}(m_{x_i}, \sigma_{x_i}^2)$. The next result immensely facilitates the computation of the quantiles of the thin-plate transport spline.

**Proposition 15.** Consider:

- $(m_x)_{x \in \mathbb{R}^2}$, the (Euclidean) thin-plate spline interpolating the means $(m_{x_i})_{i=0}^N$, and
- $(s_x)_{x \in \mathbb{R}^2}$, the (Euclidean) thin-plate spline interpolating the standard deviations $(\sigma_{x_i})_{i=0}^N$.

For any $\alpha \in [0, 1]$, the $\alpha$-quantile of $\mu_x$, the interpolated thin-plate transport spline at $x$, is given by $m_x + \Phi^{-1}(\alpha) |s_x|$.

**Proof.** It is standard that there is a linear mapping $S_x$ such that the Euclidean thin-plate spline interpolating through $(x_i, z_i)_{i=0}^N$ is given by $S_x(z_0, z_1, \ldots, z_N)$.

---

10 The inverse CDFs described here exist because of our assumption of absolute continuity of the measures.
It follows from (5.1) and the discussion in Appendix F.1 that the Monge map from $\mu^*_{x_0}$ to $\mu^*_{x_i}$ is the increasing map $z \mapsto \left(\sigma_{x_i}/\sigma_{x_0}\right)(z - m_{x_0}) + m_{x_i}$. Thus,

$$X_x = S_x(X_{x_0}, X_{x_1}, \ldots, X_{x_N})$$

$$= S_x(X_{x_0}, \frac{\sigma_{x_1}}{\sigma_{x_0}}(X_{x_0} - m_{x_0}) + m_{x_1}, \ldots, \frac{\sigma_{x_N}}{\sigma_{x_0}}(X_{x_0} - m_{x_0}) + m_{x_N})$$

$$= S_x(m_{x_0}, m_{x_1}, \ldots, m_{x_N}) + \frac{X_{x_0} - m_{x_0}}{\sigma_{x_0}} S_x(\sigma_{x_0}, \sigma_{x_1}, \ldots, \sigma_{x_N})$$

$$= m_x + \frac{X_{x_0} - m_{x_0}}{\sigma_{x_0}} S_x(\sigma_{x_0}, \sigma_{x_1}, \ldots, \sigma_{x_N})$$

$$= m_x + s_x \frac{X_{x_0} - m_{x_0}}{\sigma_{x_0}} \sim \mathcal{N}(m_x, s^2_x) = \mu_x.$$  

This is the desired result. \hfill \blacksquare

F.3 Figure 4

Here we give more details on the thin-plate spline interpolation leading to Figure 4. The data is a representation of the temperature at various weather stations throughout California on June 1 of each year in a thirty year period. That is, we consider the distribution of temperatures recorded on each of June 1, 1981, June 1, 1982, …, June 1, 2010, and we model this distribution as Gaussian (characterized by its mean and standard deviation). This data is processed and released each decade by the NOAA NCEI (Arguez et al., 2010). We interpolate these measures using our transport spline technique, obtaining Gaussian measures at each point in California. The left side of Figure 4 summarizes these measures by their quantiles, while the right side illustrates the behavior of our method as we sample increasingly many weather stations. The median temperature in the top left quantile plot is taken to be equal to the mean temperature due to our assumption that the temperature distribution is Gaussian at every location. Though there are 484 stations in the NOAA dataset, we used substantially fewer to better capture the convergence of our method.

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.

Arguez, Anthony et al. (2010). NOAA’s U.S. Climate Normals (1981-2010) [Daily]. NOAA National Centers for Environmental Information. doi: 10.7289/V5PN93JP. (Visited on 10/14/2020).

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. 1–26.

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.

Carmo, Manfredo P. do (1992). Riemannian geometry. Mathematics: Theory & Applications. Translated from the second Portuguese edition by Francis Flaherty. Birkhäuser Boston, Inc., Boston, MA, pp. xiv+300.

– (2016). Differential geometry of curves & surfaces. Dover Publications, Inc., Mineola, NY, pp. xvi+510.

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 (P2(M), W2)”. 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.

Hall, Charles A. and Meyer, W. Weston (1976). “Optimal error bounds for cubic spline interpolation”. In: J. Approximation Theory 16.2, pp. 105–122.

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.

Lehmann, Erich L. and Romano, Joseph P. (2005). Testing statistical hypotheses. Third. Springer Texts in Statistics. Springer, New York, pp. xiv+784.

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.

Range, R. Michael (1986). Holomorphic functions and integral representations in several complex variables. Vol. 108. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xx+386.

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.