A DETAILS FOR THE P-SPLINE EXAMPLE

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

$$\mu^\star_{t_i} := \mathcal{N}\big(0, (1-t_i)^2 + t_i^2\big), \quad i = 0, \dots, N. \tag{9}$$

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 (5) are deterministic. Indeed, consider the following.

Proposition 4. *Let $\mu^\star_0$ and $\mu^\star_1$ be any probability measures. Then,* any *coupling (X_0, X_1) of the two measures induces an optimal P-spline solution (X_t) to (5) with data $\mu^\star_0$ and $\mu^\star_1$.*

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 (5). □

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 μ_0 to μ_1, and $X^\star_1$ is not guaranteed to be a deterministic function of $X^\star_0$. A slight modification of this simple example yields:

Proposition 5. *Let $\mu^\star_0$ be any absolutely continuous measure. Then, there exist absolutely continuous data $(\mu^\star_{i/N})_{i=1}^N$ and an optimal solution (X_t) to the P-spline problem (5) for $\mu^\star_0, \mu^\star_{1/N}, \dots, \mu^\star_1$ such that X_1 is not a deterministic function of X_0.*

Proof. Indeed, let $T, \bar{T} : \mathbb{R}^d \to \mathbb{R}^d$ be two mappings which are $\mu^\star_0$-a.e. distinct, i.e., $T \neq \bar{T}$. Draw $X_0 \sim \mu^\star_0$. Then, we either set $X_t = (1-t)X_0 + tT(X_0)$ or else $X_t = (1-t)X_0 + t\bar{T}(X_0)$ with probability $1/2$ each (with the choice being made independently of the draw of X_0). Set $\mu^\star_{i/N} := \mathrm{law}(X_{i/N})$.

By construction, the marginals of the process (X_t) at times $0, 1/N, \dots, 1$ 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 (5).

Since T and $\bar{T}$ are distinct, X_1 is not a deterministic function of X_0. Also, the mappings T and $\bar{T}$ 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 (5) can be reduced to a multimarginal optimal transport problem involving the measures $\mu^\star_{t_0}, \mu^\star_{t_1}, \dots, \mu^\star_{t_N}$,

$$\inf_{\pi \in \Pi(\mu^\star_{t_0}, \mu^\star_{t_1}, \dots, \mu^\star_{t_N})} \int c \, \mathrm{d}\pi, \tag{10}$$

where c is a quadratic cost function. The reduction is in the following sense: if π is an optimal solution for (10), then let $(X_{t_0}, X_{t_1}, \dots, X_{t_N}) \sim \pi$, 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 (5). Any optimal solution of (5) is also of this form, having sample paths that are cubic splines.

Since the cost in the multimarginal problem (10) is quadratic, it depends only on the mean and covariance matrix of the coupling π. Suppose now that the data $(\mu^\star_{t_i})_{i=0}^N$ is Gaussian, and suppose we are given any optimal coupling π for (10). Then, we can find a *jointly Gaussian* coupling $\bar{\pi}$ of the data which has the same mean and covariance structure as π, which means $\bar{\pi}$ is also optimal for (5). The coupling $\bar{\pi}$ then induces a Gaussian process $(\bar{X}_t)$ which is optimal for (5). Such a solution has the appealing property that the law μ_t of $\bar{X}_t$ is also Gaussian for every time t.

From this discussion, it is natural to restrict ourselves to solutions to (5) which are Gaussian processes. We call such a solution a *Gaussian solution* to the P-spline problem (5). We now state a counterexample which proves Proposition 1.

Proposition 6. *Assume $N > 1$. For $i = 0, 1, \dots, N$, let $\mu^\star_{t_i} = \mathcal{N}(0, (1-t_i)^2 + t_i^2)$. Then there is a unique Gaussian solution to the P-spline problem (5) and it is* not *induced by a deterministic map.*

Proof. The key observation is that the marginals $\mu^\star_{t_i}$ arise from the curve of measures formed as the law of $X^\star_t := (1-t)X^\star_0 + tX^\star_1$ for independent standard Gaussians $X^\star_0$ and $X^\star_1$. If we consider the distribution on paths which is the law of $(X^\star_t)$, then it is supported on straight lines traversed at constant speed and so it must be optimal for the P-spline problem (5), having zero objective value.

Consider some other stochastic process (X_t) such that the law of $(X_{t_i})_{i=0}^N$ is jointly Gaussian. For (X_t) to be an optimal solution to the P-spline problem (5), 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 $\mathbb{E}[X_0^2] = \mathbb{E}[X_1^2] = 1$ and so long as $N > 1$, for $i = 1, \dots, N-1$, it holds that $t_i \notin \{0, 1\}$ and

$$\begin{aligned}(1-t_i)^2 + t_i^2 &= \mathbb{E}\big[\big((1-t_i)X_0 + t_i X_1\big)^2\big] \\ &= (1-t_i)^2 + t_i^2 + 2t_i\,(1-t_i)\,\mathbb{E}[X_0 X_1].\end{aligned}$$

Therefore $\mathbb{E}[X_0 X_1] = 0$ and (X_t) has the same distribution as $(X^\star_t)$. Consequently, the unique jointly Gaussian solution to the P-spline problem is $(X^\star_t)$. Clearly, the path $(X^\star_t)$ is not a deterministic function of $X^\star_0$. Indeed, $X^\star_1$ is *independent* of $X^\star_0$. □

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:

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

Denote by $E[\gamma] = \int_0^1 \|\ddot\gamma(t)\|^2\,\mathrm{d}t$ 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 γ_K.

Proposition 7. *Let H be a Hilbert space, and let $L \subseteq H$ be a closed linear subspace. Take points $x_0, \dots, x_N \in L$. Then the solution γ_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 γ interpolates the points $(x_i)_{i=0}^N$. Then for any admissible curve $\gamma(t) = P\gamma(t) + (I-P)\gamma(t)$, so $\ddot\gamma(t) = P\ddot\gamma(t) + (I-P)\ddot\gamma(t)$ as well. Since these two terms are orthogonal, we have

$$\|\ddot\gamma(t)\|^2 = \|P\ddot\gamma(t)\|^2 + \|(I-P)\ddot\gamma(t)\|^2.$$

Thus, on the one hand, if $\bar\gamma(t) = P\gamma_H(t)$ then $E[\bar\gamma] \le E[\gamma_H]$, and $\bar\gamma$ is interpolating because $x_i \in L$. On the other hand, $E[\gamma_H] \le E[\gamma_L] \le E[\bar\gamma]$ 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 $x_1, \dots, x_n \in K$. 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 $f\colon [0,1] \to H$ 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 γ_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 γ_K we then have $E[\gamma_K + \varepsilon f] \geq E[\gamma_K]$. Thus γ_K is stationary for E considered on H, and because E is strictly convex it follows that γ_K is optimal for (E_H) and is therefore equal to γ_H by uniqueness. □

Proposition 9. *Let $\mu^\star_{t_0}, \mu^\star_{t_1}, \dots, \mu^\star_{t_N}$ be Gaussian measures on $\mathbb{R}$. Consider the Gaussian version of the E-spline problem on $\mathbb{R}$:*

$$\min_{(\gamma_t)} \int_0^1 \|\nabla_{v_t} v_t\|^2_{L^2(\gamma_t)} \, \mathrm{d}t \quad \textit{s.t.} \quad \gamma_{t_i} = \mu^\star_{t_i},\ i = 1, \dots, N$$

where the minimization is taken over curves (γ_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 (4) in $\mathcal{P}_2(\mathbb{R})$ with the added constraint that the measures are Gaussian. If there is an optimal solution $(\gamma^\star_t)$ which is a non-degenerate Gaussian for all time, then it is also the solution to the E-spline problem (4).

Proof. It is known that $\mathcal{P}_2(\mathbb{R})$ 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 μ. 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 $\mathcal{P}_2(\mathbb{R})$, and it has closed span because it is finite-dimensional. In light of this isometry the E-spline problem (4) is equivalent to (E_S) while the Gaussian E-spline problem stated in the proposition is equivalent to (E_K) and $\gamma^\star = \gamma_K$ (the preservation of E-splines under isometry is discussed in Appendix A.3).

Applying Proposition 8 to $\gamma^\star = \gamma_K$, we deduce that $\gamma^\star = \gamma_H$. Moreover, $E[\gamma_H] \leq E[\gamma_S] \leq E[\gamma^\star]$, whence by uniqueness we get that $\gamma^\star = \gamma_S$ 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 (μ_t) be a P-spline with initial and final data μ_0 and μ_1 which are Gaussian, and assume:*

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

Then (μ_t) is induced by a jointly Gaussian coupling of μ_0 and μ_1.

Proof. Since (μ_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 (μ_t) coupling, then

$$X_t = (1-t)X_0 + tX_1 \tag{11}$$

and by assumption this variable is Gaussian. Let Z be the Gaussian with the same covariance structure as X. Scaling (11) by a positive constant, we get, for all $a, b \geq 0$

$$\langle (a,b), X \rangle \overset{\mathrm{d}}{=} \langle (a,b), Z \rangle$$

where we mean equality in distribution. This implies

$$\varphi_X(a,b) = \varphi_Z(a,b)$$

where φ_Y denotes the characteristic function of Y and is defined by $\varphi_Y(z) = \mathbb{E}[e^{i\langle z, Y\rangle}]$. Now, it is well-known that if $\mathbb{E}e^{m\|Y\|} < \infty$ for some $m > 0$ then φ_Y continues to a holomorphic function in the strip $\{z \mid |\mathrm{Im}\, z_i| < m \ \forall i\}$ (Lehmann and Romano, 2005, Theorem 2.7.1). In particular, if Y has sub-Gaussian tails, φ_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*

$$\{(z_1 + x_1, \dots, z_d + x_d) \in \mathbb{C}^d \mid |\operatorname{Re} x_i| < r \text{ for } i = 1, \dots, d\}.$$

If f and g agree on a real cube of positive radius about z, then $f \equiv g$ on all of Ω.

Now, X has sub-Gaussian tails. Indeed,

$$M_X(t) = \mathbb{E}\, e^{\langle t, X\rangle} = \mathbb{E}\, e^{t_1 X_0 + t_2 X_1} \le \left(\mathbb{E}\, e^{2t_1 X_0}\ \mathbb{E}\, e^{2t_2 X_1}\right)^{1/2} = e^{t_1^2 \operatorname{var} X_0 + t_2^2 \operatorname{var} X_1}$$

where M_X denotes the moment generating function of X. Thus φ_X is entire, along with φ_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 $X \overset{\mathrm{d}}{=} Z$. Thus X is jointly Gaussian. □

Proposition 2 is implied by the following result.

Proposition 10. *For $i = 0, \dots, N$, let $\mu^\star_{t_i} = \mathcal{N}(0, \sigma^2_{t_i})$, where $\sigma_t^2 = (1-t)^2 + t^2$. Then for all $N \ge 2$, the E-spline (4) and P-spline (5) interpolations do not coincide.*

Before starting the proof, we dispense with a possible source of confusion. The solution to the P-spline problem (5) is a stochastic process (X_t); on the other hand, the E-spline solution yields a natural stochastic process, namely the Lagrangian coupling $(X^\star_t)$ (see Section 2). In the proposition, we are not asserting that the process (X_t) and $(X^\star_t)$ 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 strictly stronger statement.

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

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

Suppose we have data $\mu^\star_{t_i} = \mathcal{N}(0, \sigma_i^2)$ at times t_i and let $t \mapsto \gamma(t)$ be the Euclidean spline interpolation of $(t_i, \sigma_i)_{i=0}^N$ on $\mathbb{R}$. It is possible that $\gamma(t) \le 0$ 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

$$\mu_t^{\mathrm{E}} = \mathcal{N}\big(0, \gamma(t)^2\big), \qquad t \in [0, 1],$$

and by Proposition 9 this must coincide with the Wasserstein E-spline (4). This is all under the hypothesis that $\gamma(t) > 0$.

Now substitute our example, with $\sigma_i^2 = (1 - t_i)^2 + t_i^2$. 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

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

For $N \ge 2$ 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.

[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.

Let (μ_t^{P}) be an interpolating P-spline. It is possible that this is not unique, but if μ_t^{P} is not Gaussian for some t then we are done, since μ_t^{E} is Gaussian by Proposition 9. Applying Lemma 1, we see that μ_t^{P} must be induced by a jointly Gaussian coupling of $\mu_0^\star$ and $\mu_1^\star$, so by Proposition 6 it must be that $\mu_t^{\mathrm{P}} = \mathcal{N}(0, (1-t)^2 + t^2)$.

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

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 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.[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 (μ_t) is a curve of measures in $\mathcal{P}_2(\mathbb{R})$ and let (v_t) be the corresponding tangent vectors. The relationship between (μ_t) and (v_t) is given by the *continuity equation* (Ambrosio, Gigli, and Savaré, 2008, Theorem 8.3.1):

$$\partial_t \mu_t + (\mu_t v_t)' = 0. \tag{12}$$

Here, we use ∂_t for the time derivative, and we use $'$ to denote spatial derivatives. If F_μ denotes the CDF of μ, then (12) implies

$$\partial_t F_{\mu_t}(x) = \partial_t \int_{-\infty}^{x} \mathrm{d}\mu_t = -\int_{-\infty}^{x} (\mu_t v_t)' = -\mu_t(x) v_t(x).$$

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

$$\begin{aligned} 0 &= (\partial_t F_{\mu_t}^{-1})\big(F_{\mu_t}(x)\big) + (F_{\mu_t}^{-1})'\big(F_{\mu_t}(x)\big) \\ &= (\partial_t F_{\mu_t}^{-1})\big(F_{\mu_t}(x)\big) + \frac{1}{F'_{\mu_t}(x)} \\ &= (\partial_t F_{\mu_t}^{-1})\big(F_{\mu_t}(x)\big) + \frac{1}{\mu_t(x)}, \end{aligned}$$

where we have applied the inverse function theorem. Thus,

$$(\partial_t F_{\mu_t}^{-1})(\alpha) = v_t\big(F_{\mu_t}^{-1}(\alpha)\big). \tag{13}$$

Differentiating again,

$$\begin{aligned} (\partial_t^2 F_{\mu_t}^{-1})(\alpha) &= (\partial_t v_t)\big(F_{\mu_t}^{-1}(\alpha)\big) + v_t'\big(F_{\mu_t}^{-1}(\alpha)\big)(\partial_t F_{\mu_t}^{-1})(\alpha) \\ &= (\partial_t v_t + v_t' v_t)\big(F_{\mu_t}^{-1}(\alpha)\big). \end{aligned}$$

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

$$\begin{aligned} \int_0^1 |\partial_t^2 F_{\mu_t}^{-1}|^2 &= \int_0^1 |(\partial_t v_t + v_t' v_t) \circ F_{\mu_t}^{-1}|^2 \\ &= \int |\partial_t v_t + v_t' v_t|^2 \,\mathrm{d}\mu_t \\ &= \|\nabla_{v_t} v_t\|_{L^2(\mu_t)}^2, \end{aligned}$$

[7] The word *formal* here, meaning that the argument proceeds by manipulating the *form* of the expressions, is not a synonym for "rigorous".

[8] In fact, this is related to Gauss's famous *Theorema Egregium*, see Carmo (2016, §4.3) and Carmo (1992, Remark 2.7).

where we use the fact that the pushforward of the uniform distribution on $[0,1]$ under $F_{\mu_t}^{-1}$ is μ_t. This equation shows that the norm (measured in H) of the acceleration of the curve $t \mapsto F_{\mu_t}^{-1}$ 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 $\mathcal{P}_2(\mathbb{R}) \hookrightarrow H$.

Remark 2. From the equation (13), we can also read off the isometry between the tangent space of H and the tangent space of $\mathcal{P}_2(\mathbb{R})$.

The reader who is uncomfortable with the formal derivation above can instead use the isometric embedding $\mathcal{P}_2(\mathbb{R}) \hookrightarrow L^2[0,1]$ as the definition of the geometry of $\mathcal{P}_2(\mathbb{R})$ (and thus, the definition of E-splines on $\mathcal{P}_2(\mathbb{R})$). 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.

B E-SPLINES AND TRANSPORT SPLINES IN ONE DIMENSION

In this section, we investigate the relationship between transport splines and E-splines on $\mathcal{P}_2(\mathbb{R})$, 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 $\mu_{t_i}^\star$ are absolutely continuous in order to properly define the covariant derivative. However, the embedding $\mathcal{P}_2(\mathbb{R}) \hookrightarrow L^2[0,1]$ allows us to rigorously extend the definition of an E-spline on all of $\mathcal{P}_2(\mathbb{R})$.

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

$$X_{t_i} := F_{\mu_{t_i}^\star}^\dagger(U) \sim \mu_{t_i}^\star, \qquad i = 0, 1, \dots, N.$$

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

Since each X_{t_i} 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 $\left(F_{\mu_{t_i}^\star}^\dagger\right)_{i=0}^N$, 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 μ_t of X_t. By the characterization of Monge maps in one dimension (Appendix F.1), it follows that $G_t = F_{\mu_t}^\dagger$.

Since (G_t) is a cubic spline, then it minimizes curvature, i.e., it solves the problem

$$\inf_{(G_t)} \int_0^1 \|\ddot{G}_t\|_{L^2[0,1]}^2 \,\mathrm{d}t, \quad \text{s.t.} \quad G_{t_i} = F_{\mu_{t_i}^\star}^\dagger \text{ for all } i.$$

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

$$\inf_{(\mu_t)} \int_0^1 \|\partial_t^2 F_{\mu_t}^\dagger\|_{L^2[0,1]}^2 \,\mathrm{d}t, \quad \text{s.t.} \quad \mu_{t_i} = \mu_{t_i}^\star \text{ for all } i,$$

since the the first problem is a relaxation of the second (given a solution (μ_t) of the second problem, we can obtain a solution $(G_t) = (F_{\mu_t}^\dagger)$ 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

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

where the infimum is taken over curves (μ_t) in $\mathcal{P}_2(\mathbb{R})$ and their corresponding tangent vectors (v_t). This problem is seen to be the E-spline problem (4).

We have thus shown that (μ_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 $(X_t^\star)$ associated with the E-spline has the same law as (X_t). For this, we can simply appeal to the embedding $\mathcal{P}_2(\mathbb{R}) \hookrightarrow H$ again. Indeed, since $\dot{X}_t = \partial_t F^\dagger_{\mu_t}(U)$, the calculation in Appendix A.3 shows that $\dot{X}_t = v_t(X_t)$ where (v_t) is the tangent vector to (μ_t), so in fact (X_t) is the Lagrangian coupling of (μ_t). □

In particular, since the Gaussian measures form a 2 dimensional half-subspace of $L^2[0,1]$ with the usual identification $\mathcal{P}_2(\mathbb{R}) \hookrightarrow L^2[0,1]$, 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 $\mathcal{P}_2(\mathbb{R})$). 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 $\mathcal{P}_2(\mathbb{R}) \subset L^2[0,1]$. First, we give a simple Gaussian counterexample.

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

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

Then, the E-spline (4) interpolation (μ_t^{E}) and transport spline interpolation (μ_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 $(X_0, X_{1/3}, X_{2/3}, X_1) = (X_0, \delta X_0, \delta X_0, X_0)$ 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

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

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

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 δ. It can be checked that at time $1/3$, the curve $t \mapsto |S_t(1,\delta,\delta,1)|$ is not $\mathcal{C}^2$ 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

$$\mu_0^\star = \mu_1^\star = \text{uniform on } [-(1+\delta),-1] \cup [1,1+\delta], \qquad \mu_{1/4}^\star = \mu_{3/4}^\star = \text{uniform on } [-\delta,\delta]. \tag{14}$$

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

$$\inf_{(\mu_t)} \int_0^1 \int_0^1 \|\ddot{F}_t^\dagger(u)\|^2 \,\mathrm{d}u\,\mathrm{d}t \qquad \text{s.t.} \qquad \mu_t = \mu_t^\star \text{ for all } t \in \{0, 1/4, 3/4, 1\},$$

where $F_t^\dagger$ denotes the quantile function of μ_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

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

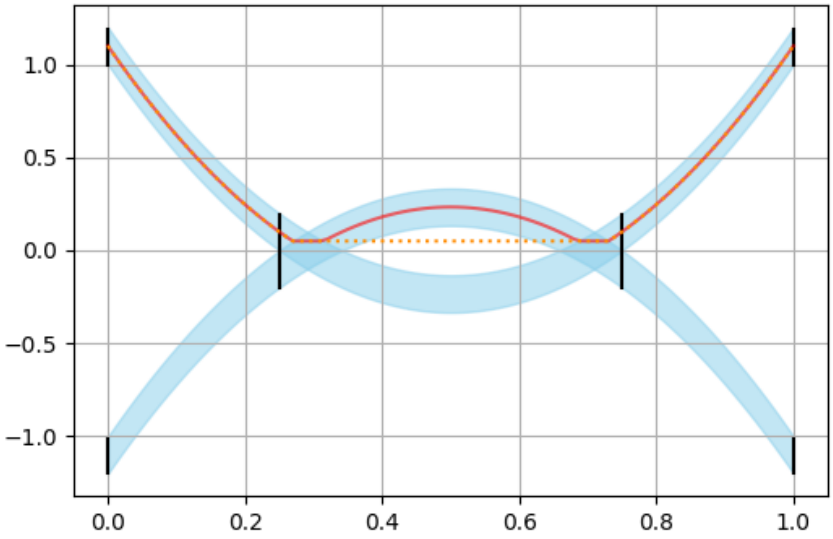

Figure 5: Transport splines interpolation for the four uniform distributions as in (14). The red line is the quantile of order 3/4 for the interpolation and the orange dotted line represents the corresponding candidate $\bar{F}_t^\dagger(u)$ for $u = 3/4$ introduced in (15).

Clearly, for δ small enough the quantiles $F_t^\dagger(u)$ 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 $\partial_t F_t^\dagger(u)|_{t=t_u^-} = \partial_t F_t^\dagger(u)|_{t=t_u^+} = 0$ and $|\partial_t^2 F_t^\dagger(u)| > 0$ for $t \in (t_u^-, t_u^+)$. One can check then that the function $u \mapsto \bar{F}_t^\dagger$ at time $t \in [0,1]$ defined by

$$\bar{F}_t^\dagger(u) = \begin{cases} F_{t_u^-}^\dagger(u), & u \in (t_u^-, t_u^+) \\ F_t^\dagger(u), & \text{otherwise} \end{cases} \tag{15}$$

is a quantile function. In particular, the measures with quantiles $\bar{F}_t^\dagger$ interpolate the measures (14) and

$$|\partial_t^2 \bar{F}_t^\dagger(u)| = \begin{cases} 0, & u \in (t_u^-, t_u^+) \\ |\partial_t^2 F_t^\dagger(u)|, & \text{otherwise,} \end{cases}$$

ensuring that $\bar{F}_t^\dagger$ 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.

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 $(X_t^\star)$ denote the Lagrangian coupling for $(\mu_t^\star)_t$, and let (X_t) be the stochastic process associated with the transport spline. Since $\mu_{t_{i-1}} = \mu_{t_{i-1}}^\star$, we can couple the two processes together so that $X_{t_{i-1}} = X_{t_{i-1}}^\star$.

By the definition of the Wasserstein distance, we can bound $W_2(\mu_t, \mu_t^\star) \le \|X_t - X_t^\star\|_{L^2(\mathbb{P})}$, so it suffices to show that the trajectories (X_t) and $(X_t^\star)$ 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

$$\sup_{t \in [0,\delta]} \{\|\ddot{x}(t)\| \vee \|\ddot{y}(t)\|\} \le R,$$

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

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

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 $(\mu_t^\star)$, we have a corresponding curve of covariance matrices (Σ_t) such that $\mu_t^\star = \mathcal{N}(0, \Sigma_t)$.

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 $X^\star \sim \mathcal{N}(0, \Sigma)$, then $\|AX^\star\|_{L^2(\mathbb{P})} = \|A\|_\Sigma$. From our eigenvalue bound we have $\|A\|_\Sigma \ge \sqrt{\lambda_{\min}(\Sigma)}\,\|A\|_{\mathrm{F}}$.

The Monge map T between two Gaussians is the linear map $T(X)$ given in (6) 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 $(v_t^\star)$ associated to the Lagrangian coupling of the curve, is also linear for each t: $v_t^\star$ is a symmetric linear mapping $\mathbb{R}^d \to \mathbb{R}^d$, that is, there exists a symmetric matrix $V_t^\star \in \mathbb{R}^{d\times d}$ such that $v_t^\star(x) = V_t^\star x$.

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 $\mathrm{id} + \delta_i v_{t_{i-1}}$. We prove a more general approximation result which may be of independent interest.

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

1. *$\bar{T}$ can be identified with a symmetric matrix.*

2. $\|\bar{T}X_t^\star - X_t^\star\|_{L^2(\mathbb{P})} \le c\sqrt{\lambda_{\min}(\Sigma_t)}$.

Then,

$$\|TX_t^\star - \bar{T}X_t^\star\|_{L^2(\mathbb{P})} \le \frac{1+2c}{1-c}\,\|\bar{T}X_t^\star - X_{t+h}^\star\|_{L^2(\mathbb{P})}.$$

Proof. Let $e := X_{t+h}^\star - \bar{T}X_t^\star$.

Consider the quadratic function $\varphi : \mathbb{R}^d \to \mathbb{R}$ 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,

$$\mathbb{E}\,\varphi(TX_t^\star) = \mathbb{E}\,\varphi(X_{t+h}^\star) = \mathbb{E}\,\varphi(\bar{T}X_t^\star + e).$$

Expanding this out,

$$\begin{aligned} 0 &= \mathbb{E}\langle (T+\bar{T})X_t^\star + e, A\{(T-\bar{T})X_t^\star - e\}\rangle \\ &= \mathbb{E}\langle (T+\bar{T})X_t^\star, A(T-\bar{T})X_t^\star\rangle + \text{error}. \end{aligned}$$

We next bound the error term. First, note that by our assumption,

$$\begin{aligned} &\|T - I_d\|_{\Sigma_t} = W_2(\mu_t^\star, \mu_{t+h}^\star) \le L\delta \le c\sqrt{\lambda_{\min}}, \\ &\|\bar{T} - I_d\|_{\Sigma_t} \le c\sqrt{\lambda_{\min}}, \end{aligned}$$

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

$$\begin{aligned} |\mathbb{E}\langle e, A(T-\bar{T})X_t^\star\rangle| &\le \|e\|_{L^2(\mathbb{P})}\,\|A(T-\bar{T})\|_{\Sigma_t} \\ &\le \|e\|_{L^2(\mathbb{P})}\,\|A\|_{\mathrm{F}}\,\|T-\bar{T}\|_{\Sigma_t} \\ &\le \|e\|_{L^2(\mathbb{P})}\,\frac{1}{\sqrt{\lambda_{\min}}}\,(\|T-I_d\|_{\Sigma_t} + \|\bar{T}-I_d\|_{\Sigma_t}) \\ &\le 2c\,\|e\|_{L^2(\mathbb{P})}, \end{aligned}$$

where we used the fact that $\|A\|_{\Sigma_t} \le 1$ implies that $\|A\|_{\mathrm{F}} \le 1/\sqrt{\lambda_{\min}}$. The second term is bounded by

$$\begin{aligned} |\mathbb{E}\langle (T+\bar{T})X_t^\star + e, Ae\rangle| &\le |\mathbb{E}\langle TX_t^\star + X_{t+h}^\star - 2X_t^\star, Ae\rangle| + 2|\mathbb{E}\langle X_t^\star, Ae\rangle| \\ &\le \{\|A\|_{\mathrm{F}}\,(\|T-I_d\|_{\Sigma_t} + \|X_{t+h}^\star - X_t^\star\|_{L^2(\mathbb{P})}) + 2\|A\|_{\Sigma_t}\}\,\|e\|_{L^2(\mathbb{P})} \\ &\le 2\,(1+c)\,\|e\|_{L^2(\mathbb{P})}, \end{aligned}$$

where we used

$$\|X_{t+h}^\star - X_t^\star\|_{L^2(\mathbb{P})}^2 = \mathbb{E}\Big[\Big\|\int_t^{t+h} \dot{X}_s^\star\,\mathrm{d}s\Big\|^2\Big] \le \delta\,\Big|\int_t^{t+h} \|\dot{X}_s^\star\|_{L^2(\mathbb{P})}^2\,\mathrm{d}s\Big| \le L^2\delta^2.$$

Thus, we have

$$\begin{aligned} 2\|T-\bar{T}\|_{\Sigma_t} &= 2\,\mathbb{E}\langle X_t^\star, A(T-\bar{T})X_t^\star\rangle \\ &= -\,\mathbb{E}\langle (T+\bar{T}-2I_d)X_t^\star, A(T-\bar{T})X_t^\star\rangle + \text{error} \\ &\le (\|T-I_d\|_{\Sigma_t} + \|\bar{T}-I_d\|_{\Sigma_t})\,\|A\|_{\mathrm{F}}\,\|T-\bar{T}\|_{\Sigma_t} + \text{error} \\ &\le 2c\,\|T-\bar{T}\|_{\Sigma_t} + (2+4c)\,\|e\|_{L^2(\mathbb{P})} \end{aligned}$$

which finally yields

$$\|T-\bar{T}\|_{\Sigma_t} \le \frac{1+2c}{1-c}\,\|e\|_{L^2(\mathbb{P})}.$$

□

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 δ is small enough so that*

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

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

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

Proof. We apply Theorem 3 with

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

Using $\dot{X}_t^\star = V_t^\star X_t^\star$, where $V_t^\star$ is symmetric, we obtain:

$$\begin{aligned}
\dot{X}_t^\star &= V_t^\star X_t^\star, \\
\ddot{X}_t^\star &= \dot{V}_t^\star X_t^\star + V_t^{\star 2} X_t^\star = (\dot{V}_t^\star + V_t^{\star 2}) X_t^\star, \\
\dddot{X}_t^\star &= (\ddot{V}_t^\star + 2\dot{V}_t^\star V_t^\star + V_t^\star \dot{V}_t^\star + V_t^{\star 3}) X_t^\star, \\
&\vdots
\end{aligned}$$

Observe that the ith derivative of $t \mapsto X_t^\star$ at t is indeed a linear function of $X_t^\star$, 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

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

by our assumption on δ.

Finally, the error $e := X_{t+h}^\star - \bar{T} X_t^\star$ is controlled via Taylor's theorem:

$$\begin{aligned}
\|e\|_{L^2(\mathbb{P})} &= \left\| X_{t+h}^\star - \sum_{i=0}^{k} \frac{h^i}{i!} (\partial^i X^\star)_t \right\|_{L^2(\mathbb{P})} \\
&= \left\| \int_t^{t+h} \frac{(\partial^{k+1} X^\star)_s}{k!} (s-t)^k \, \mathrm{d}s \right\|_{L^2(\mathbb{P})} \\
&\leq \frac{R_{k+1}\delta^{k+1}}{(k+1)!}.
\end{aligned}$$

□

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

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

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

Corollary 1 says that in order to prove our desired result $\dot{X}_{t_{i-1}} = \dot{X}_{t_{i-1}}^\star + O(\delta)$, it suffices to show that $\dot{X}_{t_{i-1}} = (T_i X_{t_{i-1}} - X_{t_{i-1}})/\delta_i + O(\delta)$ (since the RHS of both expressions equals $V_{t_{i-1}}^\star X_{t_{i-1}} = V_{t_{i-1}}^\star X_{t_{i-1}}^\star$ 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

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

Therefore, we next control these quantities.

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

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

Proof. From Corollary 1,

$$\Big\| \frac{X_{t_i} - X_{t_{i-1}}}{\delta_i} - V^\star_{t_{i-1}} X_{t_{i-1}} \Big\|_{L^2(\mathbb{P})} = \Big\| \frac{T_i - I_d}{\delta_i} - V^\star_{t_{i-1}} \Big\|_{\Sigma_{t_{i-1}}} \le 2R\delta_i,$$

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

$$\Big\| \frac{X_{t_{i+1}} - X_{t_i}}{\delta_{i+1}} - V^\star_{t_i} X_{t_i} \Big\|_{L^2(\mathbb{P})} \le 2R\delta_{i+1}.$$

Therefore,

$$\|\Delta_i\|_{L^2(\mathbb{P})} \le 4R\delta + \|V^\star_{t_i} X_{t_i} - V^\star_{t_{i-1}} X_{t_{i-1}}\|_{L^2(\mathbb{P})}.$$

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

$$\|V^\star_{t_i} X_{t_i} - V^\star_{t_{i-1}} X_{t_{i-1}}\|_{L^2(\mathbb{P})} \le \|V^\star_{t_i}(T_i - I_d - \delta_i V^\star_{t_{i-1}}) X_{t_{i-1}}\|_{L^2(\mathbb{P})} + \|V^\star_{t_i}(I_d + \delta_i V^\star_{t_{i-1}}) X_{t_{i-1}} - V^\star_{t_{i-1}} X_{t_{i-1}}\|_{L^2(\mathbb{P})}.$$

We control the first term using Corollary 1:

$$\begin{aligned}
\|V^\star_{t_i}(T_i - I_d - \delta_i V^\star_{t_{i-1}}) X_{t_{i-1}}\|_{L^2(\mathbb{P})} &\le \|V^\star_{t_i}\|_{\mathrm{F}} \, \|(T_i - I_d - \delta_i V^\star_{t_{i-1}}) X_{t_{i-1}}\|_{L^2(\mathbb{P})} \\
&\le \frac{L}{\sqrt{\lambda_{\min}}} \, \|T_i - I_d - \delta_i V^\star_{t_{i-1}}\|_{\Sigma_{t_{i-1}}} \\
&\le \frac{L}{\sqrt{\lambda_{\min}}} \cdot 2R\delta_i^2 \le R\delta_i,
\end{aligned}$$

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

$$\begin{aligned}
&\|V^\star_{t_i}(I_d + \delta_i V^\star_{t_{i-1}}) X_{t_{i-1}} - V^\star_{t_{i-1}} X_{t_{i-1}}\|_{L^2(\mathbb{P})} \\
&\quad \le \|V^\star_{t_i} X^\star_{t_i} - V^\star_{t_{i-1}} X^\star_{t_{i-1}}\|_{L^2(\mathbb{P})} + \|V^\star_{t_i}\{(I_d + \delta_i V^\star_{t_{i-1}}) X^\star_{t_{i-1}} - X^\star_{t_i}\}\|_{L^2(\mathbb{P})}.
\end{aligned}$$

It is easy to control

$$\|V^\star_{t_i} X^\star_{t_i} - V^\star_{t_{i-1}} X^\star_{t_{i-1}}\|_{L^2(\mathbb{P})} = \Big\| \int_{t_{i-1}}^{t_i} \ddot{X}^\star_t \, \mathrm{d}t \Big\|_{L^2(\mathbb{P})} \le R\delta_i.$$

Lastly,

$$\begin{aligned}
\|V^\star_{t_i}\{(I_d + \delta_i V^\star_{t_{i-1}}) X^\star_{t_{i-1}} - X^\star_{t_i}\}\|_{L^2(\mathbb{P})} &\le \|V^\star_{t_i}\|_{\mathrm{F}} \, \|X^\star_{t_i} - X^\star_{t_{i-1}} - \delta_i V^\star_{t_{i-1}} X^\star_{t_{i-1}}\|_{L^2(\mathbb{P})} \\
&\le \frac{L}{\sqrt{\lambda_{\min}}} \Big\| \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \ddot{X}^\star_s \, \mathrm{d}s \, \mathrm{d}t \Big\|_{L^2(\mathbb{P})} \\
&\le \frac{L}{\sqrt{\lambda_{\min}}} \cdot \frac{R\delta_i^2}{2} \le \frac{R\delta_i}{4}.
\end{aligned}$$

Putting it all together, we obtain

$$\|\Delta_i\|_{L^2(\mathbb{P})} \le \frac{25}{4} R\delta.$$

□

To match notation with Appendix D, we set

$$M_i := \ddot{X}_{t_{i-1}}, \qquad i \in [N+1].$$

Lemma 2. *Assume $\delta \le \sqrt{\lambda_{\min}}/(2L)$. It holds that*

$$\|M_i\|_{L^2(\mathbb{P})} \le \frac{75(1+\alpha)^2}{4\alpha^3} R.$$

Proof. As described in Appendix D, we know that $M = 6\mathbf{T}^{-1}\Delta$, where the entries of $\mathbf{T}^{-1}$ are bounded in Lemma 3. Thus,

$$\begin{aligned}
\|M_i\|_{L^2(\mathbb{P})} &= 6 \Big\| \sum_{j=1}^{N-1} (\mathbf{T}^{-1})_{i,j} \Delta_j \Big\|_{L^2(\mathbb{P})} \\
&\le 6 \sum_{j=1}^{N-1} |(\mathbf{T}^{-1})_{i,j}| \, \|\Delta_j\|_{L^2(\mathbb{P})} \\
&\le 6 \sum_{j=1}^{N-1} \frac{1}{4\alpha^2\delta} \frac{1}{(1+\alpha)^{|i-j|-1}} \frac{25}{4} R\delta \\
&\le \frac{75R}{4\alpha^2} \sum_{k=0}^{\infty} \frac{1}{(1+\alpha)^{k-1}} = \frac{75(1+\alpha)^2}{4\alpha^3} R,
\end{aligned}$$

where we use Proposition 12. □

Finally, we are ready to state our control on the velocity of the trajectory (X_t).

Proposition 13. *Assume $\delta \le \sqrt{\lambda_{\min}}/(2L)$. Then,*

$$\|\dot{X}_{t_{i-1}} - \dot{X}^\star_{t_{i-1}}\|_{L^2(\mathbb{P})} \le \frac{16\alpha^3 + 75(1+\alpha)^2}{8\alpha^3} R\delta.$$

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,

$$\Big\|\dot{X}_{t_{i-1}} - \frac{X_{t_i} - X_{t_{i-1}}}{\delta_i}\Big\|_{L^2(\mathbb{P})} \le \frac{\|M_{i+1}\|_{L^2(\mathbb{P})} + 2\|M_i\|_{L^2(\mathbb{P})}}{6} \delta \le \frac{75(1+\alpha)^2}{8\alpha^3} R\delta,$$

by Lemma 2. Next, we recall that $X_{t_i} = T_i X_{t_{i-1}}$, and that (X_t) and $(X^\star_t)$ are coupled so that $X_{t_{i-1}} = X^\star_{t_{i-1}}$. Thus,

$$\begin{aligned}
\|\dot{X}_{t_{i-1}} - \dot{X}^\star_{t_{i-1}}\|_{L^2(\mathbb{P})} &\le \Big\|\dot{X}_{t_{i-1}} - \frac{T_i X_{t_{i-1}} - X_{t_{i-1}}}{\delta_i}\Big\|_{L^2(\mathbb{P})} + \Big\|\dot{X}^\star_{t_{i-1}} - \frac{T_i X^\star_{t_{i-1}} - X^\star_{t_{i-1}}}{\delta_i}\Big\|_{L^2(\mathbb{P})} \\
&\le \frac{75(1+\alpha)^2}{8\alpha^3} R\delta + 2R\delta,
\end{aligned}$$

where we invoke Corollary 1 again. □

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 $\delta \le \sqrt{\lambda_{\min}}/(2L)$. Then,*

$$\sup_{t\in[0,1]} \|\ddot{X}_t\|_{L^2(\mathbb{P})} \le \frac{75(1+\alpha)^2}{4\alpha^3} R.$$

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

$$\begin{aligned}
\|\ddot{X}_t\|_{L^2(\mathbb{P})} &= \Big\|\frac{t_i - t}{\delta_i}\,\ddot{X}_{t_{i-1}} + \frac{t - t_{i-1}}{\delta_i}\,\ddot{X}_{t_i}\Big\|_{L^2(\mathbb{P})} \\
&\le \frac{t_i - t}{\delta_i}\,\|\ddot{X}_{t_{i-1}}\|_{L^2(\mathbb{P})} + \frac{t - t_{i-1}}{\delta_i}\,\|\ddot{X}_{t_i}\|_{L^2(\mathbb{P})} \\
&\le \|\ddot{X}_{t_{i-1}}\|_{L^2(\mathbb{P})} \vee \|\ddot{X}_{t_i}\|_{L^2(\mathbb{P})} \\
&= \|M_i\|_{L^2(\mathbb{P})} \vee \|M_{i+1}\|_{L^2(\mathbb{P})} \\
&\le \frac{75(1+\alpha)^2}{4\alpha^3}\,R,
\end{aligned}$$

by Lemma 2. □

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 $(X_t^\star)$ be coupled with $X_{t_{i-1}} = X^\star_{t_{i-1}}$. Then,

$$\begin{aligned}
\|X_t - X_t^\star\|_{L^2(\mathbb{P})} &\le \delta_i\,\|\dot{X}_{t_{i-1}} - \dot{X}^\star_{t_{i-1}}\|_{L^2(\mathbb{P})} + \Big\|\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} (\ddot{X}_s - \ddot{X}_s^\star)\,\mathrm{d}s\,\mathrm{d}t\Big\|_{L^2(\mathbb{P})} \\
&\le \frac{16\alpha^3 + 75(1+\alpha)^2}{8\alpha^3}\,R\delta^2 + \frac{\delta^2}{2}\sup_{t\in[0,1]}\left(\|\ddot{X}_t\|_{L^2(\mathbb{P})} + \|\ddot{X}_t^\star\|_{L^2(\mathbb{P})}\right) \\
&\le \frac{10\alpha^3 + 75(1+\alpha)^2}{4\alpha^3}\,R\delta^2 \le \frac{115}{2\alpha^3}\,R\delta^2,
\end{aligned}$$

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

$$X_t = \frac{t_i - t}{t_i - t_{i-1}}\,X_{t_{i-1}} + \frac{t - t_{i-1}}{t_i - t_{i-1}}\,X_{t_i}.$$

Let μ_t denote the law of X_t.

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

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

Proof. As in Appendix C.4, we have

$$\begin{aligned}
\|X_t - X_t^\star\|_{L^2(\mathbb{P})} &\le \delta_i\,\|\dot{X}_{t_{i-1}} - \dot{X}^\star_{t_{i-1}^+}\|_{L^2(\mathbb{P})} + \Big\|\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t} \ddot{X}_s^\star\,\mathrm{d}s\,\mathrm{d}t\Big\|_{L^2(\mathbb{P})} \\
&\le 2R\delta^2 + \frac{1}{2}R\delta^2.
\end{aligned}$$

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

$$(T_i - I_d)X_{t_{i-1}} = (T_i - I_d)X^\star_{t_{i-1}},$$

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

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

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).

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 $(x_{t_0}, x_{t_1}, \dots, x_{t_N})$ in $\mathbb{R}^d$. Our goal is to construct a piecewise cubic polynomial interpolation $y : [0,1] \to \mathbb{R}^d$ which is $\mathcal{C}^2$ 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

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

Computing derivatives,

$$\begin{aligned}
x_{t_{i-1}} = y_i(t_{i-1}) &= d_i, \\
x_{t_i} = y_i(t_i) &= a_i \delta_i^3 + \frac{m_i}{2} \delta_i^2 + c_i \delta_i + d_i, \\
\dot{y}_i(t_{i-1}) &= c_i, \\
\dot{y}_i(t_i) &= 3 a_i \delta_i^2 + m_i \delta_i + c_i, \\
\ddot{y}_i(t_{i-1}) &= m_i, \\
\ddot{y}_i(t_i) &= 6 a_i \delta_i + m_i,
\end{aligned}$$

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:

$$\begin{aligned}
a_i &= \frac{m_{i+1} - m_i}{6\delta_i}, \\
b_i &= \frac{m_i}{2}, \\
c_i &= \frac{x_{t_i} - x_{t_{i-1}}}{\delta_i} - \frac{m_{i+1} + 2m_i}{6} \delta_i, \\
d_i &= x_{t_{i-1}}.
\end{aligned}$$

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

$$6\Delta_i = \delta_i m_i + 2(\delta_i + \delta_{i+1}) m_{i+1} + \delta_{i+2} m_{i+2}, \qquad i = 1, \dots, N-1,$$

where we have defined the quantities

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

a proxy for the second derivative of the data points.

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

$$\underbrace{\begin{bmatrix} 2(\delta_1+\delta_2) & \delta_2 & & \\ \delta_2 & \ddots & \ddots & \\ & \ddots & \ddots & \delta_{N-1} \\ & & \delta_{N-1} & 2(\delta_{N-1}+\delta_N) \end{bmatrix}}_{:=\mathbf{T}} m = 6\Delta.$$

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

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

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

Proof. We write $\mathbf{T} = \mathbf{B} + \mathbf{D}$, where

$$\mathbf{B} := \begin{bmatrix} 0 & \delta_2 & & \\ \delta_2 & \ddots & \ddots & \\ & \ddots & \ddots & \delta_{N-1} \\ & & \delta_{N-1} & 0 \end{bmatrix},$$
$$\mathbf{D} := 2\,\mathrm{diag}(\delta_1+\delta_2, \ldots, \delta_{N-1}+\delta_N).$$

Therefore,

$$\begin{aligned} \mathbf{T}^{-1} &= (\mathbf{B}+\mathbf{D})^{-1} \\ &= \mathbf{D}^{-1/2}(I_{N-1} + \mathbf{D}^{-1/2}\mathbf{B}\mathbf{D}^{-1/2})^{-1}\mathbf{D}^{-1/2} \\ &= \sum_{k=0}^{\infty}(-1)^k \mathbf{D}^{-1/2}(\underbrace{\mathbf{D}^{-1/2}\mathbf{B}\mathbf{D}^{-1/2}}_{:=\mathbf{M}})^k\mathbf{D}^{-1/2}. \end{aligned}$$

The matrix $\mathbf{M}$ is

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

where we set

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

Since $\mathbf{M}$ has non-negative entries, we have the entrywise bound

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

[9] To be precise, we should write this as the block matrix equation $(\mathbf{T}\otimes I_d)m = 6\Delta$.

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

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

Therefore,

$$\begin{aligned} |(\mathbf{T}^{-1})_{i,j}| &\le \sum_{k=0}^{\infty} \frac{1}{2\sqrt{(\delta_i + \delta_{i+1})(\delta_j + \delta_{j+1})}} \frac{\mathbb{1}_{|i-j| \le k}}{(1+\alpha)^k} \\ &\le \frac{1}{4\alpha\delta} \sum_{k=|i-j|}^{\infty} \frac{1}{(1+\alpha)^k} = \frac{1}{4\alpha^2 \, (1+\alpha)^{|i-j|-1}} \frac{1}{\delta}. \end{aligned}$$

□

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

$$\mathcal{N}\Big(\begin{bmatrix} 7(k-1) \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}\Big) \qquad \text{for } k \text{ odd}$$

and

$$\mathcal{N}\Big(\begin{bmatrix} 7(k-1) \\ 7 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\Big) \qquad \text{for } k \text{ even},$$

where $k = 1, \dots, 5$.

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

$$T(x) = \mu_2 + A(x - \mu_1), \quad A = \Sigma_1^{-\frac{1}{2}} (\Sigma_1^{\frac{1}{2}} \Sigma_2 \Sigma_1^{\frac{1}{2}})^{\frac{1}{2}} \Sigma_1^{-\frac{1}{2}}.$$

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 5×10^9 and radius 1. Each body was initialized with a position x and a velocity v drawn randomly according to

$$x \sim \mathcal{N}\Big(\begin{bmatrix} 100 \\ 100 \end{bmatrix}, \begin{bmatrix} 30 & 0 \\ 0 & 20 \end{bmatrix}\Big), \qquad v \sim \mathcal{N}\Big(\begin{bmatrix} 10 \\ -20 \end{bmatrix}, \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}\Big).$$

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.

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

$$X_{x_i} = F^{-1}_{\mu^\star_{x_i}}(U), \qquad i = 0, 1, \ldots, N.$$

Then, $(X_{x_0}, X_{x_1}, \ldots, X_{x_N})$ is a simultaneously optimal coupling of the measures $\mu^\star_{x_0}, \mu^\star_{x_1}, \ldots, \mu^\star_{x_N}$. 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 $\mu^\star_{x_i}$ to $\mu^\star_{x_j}$ is characterized as the ($\mu^\star_{x_i}$-a.e.) unique mapping which both pushes $\mu^\star_{x_i}$ forward to $\mu^\star_{x_j}$ 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 $F^{-1}_{\mu^\star_{x_j}} \circ F_{\mu^\star_{x_i}}$ satisfies these properties, and thus[10]

$$T_{i,j} = F^{-1}_{\mu^\star_{x_j}} \circ F_{\mu^\star_{x_i}}.$$

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

Suppose the measures $\mu^\star_{x_i}$, $i = 0, 1, \ldots, N$, are all one-dimensional Gaussians, and write $\mu^\star_{x_i} = \mathcal{N}(m_{x_i}, \sigma^2_{x_i})$. 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 α-quantile of μ_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 (6) and the discussion in Appendix F.1 that the Monge map from $\mu^\star_{x_0}$ to $\mu^\star_{x_i}$ is the increasing map $z \mapsto (\sigma_{x_i}/\sigma_{x_0})(z - m_{x_0}) + m_{x_i}$. Thus,

$$\begin{aligned}
X_x &= S_x(X_{x_0}, X_{x_1}, \dots, X_{x_N}) \\
&= S_x\big(X_{x_0}, \frac{\sigma_{x_1}}{\sigma_{x_0}}(X_{x_0} - m_{x_0}) + m_{x_1}, \dots, \frac{\sigma_{x_N}}{\sigma_{x_0}}(X_{x_0} - m_{x_0}) + m_{x_N}\big) \\
&= S_x(m_{x_0}, m_{x_1}, \dots, m_{x_N}) + S_x\big(X_{x_0} - m_{x_0}, \frac{\sigma_{x_1}}{\sigma_{x_0}}(X_{x_0} - m_{x_0}), \dots, \frac{\sigma_{x_N}}{\sigma_{x_0}}(X_{x_0} - m_{x_0})\big) \\
&= m_x + \frac{X_{x_0} - m_{x_0}}{\sigma_{x_0}} S_x(\sigma_{x_0}, \sigma_{x_1}, \dots, \sigma_{x_N}) \\
&= m_x + s_x \frac{X_{x_0} - m_{x_0}}{\sigma_{x_0}} \sim \mathcal{N}(m_x, s_x^2) = \mu_x.
\end{aligned}$$

This is the desired result. □

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

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).

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.

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.

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.

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

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

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

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.

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.

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