



HAL
open science

Second order models for optimal transport and cubic splines on the Wasserstein space

Jean-David Benamou, Thomas Gallouët, François-Xavier Vialard

► **To cite this version:**

Jean-David Benamou, Thomas Gallouët, François-Xavier Vialard. Second order models for optimal transport and cubic splines on the Wasserstein space. 2018. hal-01682107v1

HAL Id: hal-01682107

<https://hal.science/hal-01682107v1>

Preprint submitted on 12 Jan 2018 (v1), last revised 25 Jul 2018 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SECOND ORDER MODELS FOR OPTIMAL TRANSPORT AND CUBIC SPLINES ON THE WASSERSTEIN SPACE

JEAN-DAVID BENAMOU, THOMAS O. GALLOUËT, AND FRANÇOIS-XAVIER VIALARD

ABSTRACT. On the space of probability densities, we extend the Wasserstein geodesics to the case of higher-order interpolation such as cubic spline interpolation. After presenting the natural extension of cubic splines to the Wasserstein space, we propose simpler approach, similarly to Brenier's generalized Euler solutions. Our method is based on the relaxation of the variational problem on the path space. We propose an efficient implementation based on multimarginal optimal transport and entropic regularization in 1D and 2D. Our framework also enables extrapolation of Wasserstein geodesics via a natural convex relaxation.

1. INTRODUCTION

In this paper, we propose a variational method to generalize cubic splines on the space of densities using multimarginal optimal transport. In short, the proposed method consists in minimizing, on the space of measures on the path space, under marginal constraints, the norm squared of the acceleration. In this setting, entropic regularization can be applied and give an efficient algorithms to approximate the solution. We showcase our methodology on 1D and 2D data.

In the past few years, higher-order interpolations methods have been investigated for applications in computer vision or medical imaging, for time-sequence interpolation or regression. The most usual setting is when data are modeled as shapes, which object embedded in the Euclidean space with no preferred parametrization: space of unparametrized curves or surfaces, or images are some of the most important examples. These examples are infinite dimensional but the finite dimensional case was interesting for camera motion interpolation as first introduced in [17] and further developed in [4, 6]. Motivated by sometimes different applications, the problem of interpolation between two shapes is usually treated via the use of a Riemannian metric on the space of shapes and computing a geodesic between the two shapes. From a mathematical point of view, shape spaces are often infinite dimensional and thus, non-trivial analytical questions arise such as existence of minimizing geodesics or global well-posedness of the initial value problem associated with geodesics. A finite dimensional approximation is still possible such as in [23], in which spline interpolation is proposed for a diffeomorphic group action on a finite dimensional manifold. It has been extended for invariant higher-order lagrangians in [9, 10] on a group, still finite dimensional. A numerical implementation of the variational and shooting splines has been developed in [21] with applications to medical imaging. The question of existence of an extremum is not addressed in these publications. An attempt is given in [22] where the exact relaxation of the problem is computed in the case of the group of diffeomorphisms of the unit interval. In a similar direction, in [11] discuss the convergence of the discretization of cubic splines in some particular infinite dimensional Riemannian context on the space of shapes.

As a shape space, we are, in this paper, interested in probability densities or more generally densities endowed with the Wasserstein metric. Since the Wasserstein metric shares some similarities with a Riemannian metric on this space of probability densities, it is natural to study further higher-order models in this context. Our motivation is to answer the following practical question, what is a natural extension of cubic splines to the Wasserstein space. To the best of our knowledge, this question has not been yet addressed in the literature on optimal transport.

We present in Section 2 the notion of cubic splines on a Riemannian manifold and detail its variational formulation in Hamiltonian coordinates. We then discuss independently in Section 3 a

geometric approach to the Wasserstein space that will be useful for the introduction of our proposed method detailed in Section 4. Finally in Sections 5 and 6 we present the numerical entropic relaxation method and a possible alternative numerical method based on semi-discrete optimal transport.

2. CUBIC SPLINES ON RIEMANNIAN MANIFOLDS

Variational cubic splines on a Riemannian manifold are the minimizers of the acceleration, denoting $\frac{D}{Dt}$ the covariant derivative,

$$(2.1) \quad \mathcal{E}(x) = \int_0^1 g(x) \left(\frac{D}{Dt} \dot{x}, \frac{D}{Dt} \dot{x} \right) dt,$$

subject to constraint on the path such as constraints on the tangent space (1), $(x(i), \dot{x}(i))$ are prescribed for $i = 0, 1$, or constraints on the positions (2) such as $x(t_i) = x_i$ for a collection of times $t_i \in [0, 1]$. Under mild conditions on the constraints, on a complete Riemannian manifolds, minimizers exist, for instance in the case (2) mentioned above. The Euler-Lagrange equation associated to the functional \mathcal{E} is

$$(2.2) \quad \frac{D^3}{Dt^3} \dot{x} - R \left(\dot{x}, \frac{D}{Dt} \dot{x} \right) \dot{x} = 0,$$

where R is the curvature tensor of the Riemannian manifold M . Note that this equation is similar to a Jacobi field equation.

We now formulate the variational problem in Hamiltonian coordinates. In the finite dimensional case of a riemannian manifold, the geodesic equations are given in coordinates by

$$(2.3) \quad \frac{D}{Dt} \dot{c} = \ddot{x} + \Gamma_x(\dot{x}, \dot{x}) = 0.$$

On the other hand, in the Hamiltonian formulation, we have with $\dot{q} = k(q)p$ where $k(q)^{-1}$ is the metric at point q ,

$$(2.4) \quad \dot{p} + \partial_q H = 0$$

to encode the geodesic equation (2.3). From the two equivalent formulations (2.3) and (2.4), it can be shown that $K(q)(\dot{p} + \partial_q H) = \nabla_{\dot{c}} \dot{c}$. Therefore, it shows that the variational spline problem can be rewritten in Hamiltonian coordinates as follows

$$\int_0^1 k(x)(u, u) dt,$$

under the constraint

$$\begin{cases} \dot{x} = k(x)p \\ \dot{p} = -\partial_x H(p, x) + u, \end{cases}$$

where $H(p, x) = \frac{1}{2} \langle p, k(x)p \rangle$ and $k(x)$ is the cometric at point x . It is natural to ask whether such variational problems carry over in infinite dimensional situations such as the Wasserstein space, which will be discussed in the rest of the paper.

3. A FORMAL APPLICATION OF SPLINE INTERPOLATION TO THE WASSERSTEIN SPACE

It is well known that the Hamiltonian formulation of geodesics on the Wasserstein space, define over a riemannian manifold M , are

$$(3.1) \quad \begin{cases} \dot{\rho} + \nabla \cdot (\rho \nabla p) = 0 \\ \dot{p} + \frac{1}{2} |\nabla p|^2 = 0, \end{cases}$$

which are valid when working with smooth densities. The Hamiltonian is the following,

$$(3.2) \quad H(p, \rho) = \frac{1}{2} \int_M |\nabla p|^2 \rho d\mu_0,$$

where μ_0 is a reference measure on M .

Remark 1. Taking the gradient of the equation governing p , and denoting $v = \nabla p$, we get Burger's equation:

$$(3.3) \quad \dot{v} + (v, \nabla)v = 0,$$

where in coordinates, the operator (v, ∇) is defined as $(v, \nabla)w \doteq \sum_{i=1}^n v_i \nabla w_i$ where v, w are vector fields and n is the dimension of the M . In Lagrangian coordinates, this equation implies that

$$(3.4) \quad \ddot{\varphi} = 0,$$

if φ is the Lagrangian flow associated with the time dependent vector field v , which is well-defined under sufficient regularity conditions.

Remark 2. For the Wasserstein case, the operator is given by $k(n)p = -\nabla \cdot [n\nabla p]$ so that the (formal) computation of the covariant derivative $\nabla_{\dot{n}}\dot{n}$ on the Wasserstein space is:

$$(3.5) \quad \nabla_{\dot{n}}\dot{n} = -\nabla \cdot [n(v + (v, \nabla)v)],$$

where v is the horizontal lift associated with \dot{n} . This is proven rigorously in [15].

From a control viewpoint, we aim at minimizing $\frac{1}{2} \int_0^1 H(u, \rho) dt$ for the control system:

$$(3.6) \quad \begin{cases} \dot{\rho} + \nabla \cdot (\rho \nabla p) = 0 \\ \dot{p} + \frac{1}{2} |\nabla p|^2 = u. \end{cases}$$

Alternatively, in terms of the variables (p, q) , this amounts to minimize

$$(3.7) \quad \int_0^1 \int_M |\nabla[\dot{p} + \frac{1}{2} |\nabla p|^2]|^2 \rho \, d\mu_0 \, dt,$$

under the continuity equation constraint $\dot{\rho} + \nabla \cdot (\rho \nabla p) = 0$. It is a nonconvex optimization problem in the couple (ρ, p) . However, the key issue here is that the variational problem itself is a priori not well-posed since our formulation is valid in a smooth setting and to make it rigorous on the space of measures, we would need to compute the relaxation of this problem.

4. A SIMPLER MODEL

We first recall in this paragraph that the Wasserstein space can be seen, at least formally, as a homogeneous space as described in [12, Appendix 5] and originally in [18]. Consider the group of (smooth) diffeomorphisms of M a closed manifold, $\text{Diff}(M)$, and the space of (smooth) probability densities $\text{Dens}(M)$. The space of densities is endowed with a $\text{Diff}(M)$ action defined by the pushforward, that is to a given $\varphi \in \text{Diff}(M)$ and $\rho \in \text{Dens}(M)$, the pushforward of ρ by φ is $\text{Jac}(\varphi^{-1})\rho \circ \varphi^{-1}$. By Moser's lemma, this action is transitive, thus making the space of densities as a homogeneous space. More importantly, there exist compatible Riemannian structure between $\text{Diff}(M)$ and $\text{Dens}(M)$. Once having chosen a reference density μ_0 , the $L^2(M, \mu_0)$ metric on the diffeomorphism group descends to the Wasserstein L^2 metric on the space of densities, or in other words, the pushforward action $\varphi \mapsto \varphi_*\mu_0$ is a Riemannian submersion. An important property of Riemannian submersion is that geodesics on $\text{Dens}(M)$ are in correspondence with geodesics on the group, given by horizontal lift. This property is actually contained in Brenier's polar factorization theorem, which shows that the horizontal lift is the gradient of a convex function. The first order variational calculus is now quite well understood in the Wasserstein space, as well as some higher-order geometrical quantities such as curvature.

However, for higher-order variational problems, this reduction does not hold in general even if the Riemannian submersion structure is present. It means that a variational problem on the group can lead to a different model. More precisely, instead of minimizing the kinetic energy of the map, one can minimize the acceleration of the map, that is

Definition 1 (Monge formulation). Let $0 = t_0 < \dots < t_n = 1$, $n \geq 2$ and $\rho_1, \dots, \rho_n \in \text{Dens}(M)$ be n probability measures on M .

Minimize, among time dependent maps $\varphi(t) : M \mapsto M$,

$$(4.1) \quad \int_0^1 \int_M |\dot{\varphi}|^2 d\mu_0 dt,$$

under the marginal constraints $\varphi(t_i)_*\mu_0 = \rho_i$. This minimizing problem is denoted by (MP) .

It is a Monge formulation of the variational problem, similar to standard optimal transport. On a Riemannian manifold M , the notation $\dot{\varphi}$ stands for $\frac{D}{Dt}\varphi$. By the change of variable with the map φ , the problem can be written in Eulerian coordinates, that is using the vector field associated with the Lagrangian map φ , $\partial_t\varphi = v \circ \varphi$, one aims at minimizing

$$(4.2) \quad \int_0^1 \int_M |u|^2 \rho d\mu_0 dt$$

under the constraints

$$(4.3) \quad \begin{cases} \dot{\rho} + \operatorname{div}(\rho v) = 0 \\ \dot{v} + (v, \nabla)v = u, \end{cases}$$

and some marginals constraints such as $\rho(t_i) = \rho_i$.

Remark 1. *It is now very clear that the two models (3.7) and (4.1) do not coincide a priori, since the optimization set of the first model (3.7) is a subset of the second (4.1). Indeed, if one takes $v = \nabla p$, the two formulations coincide.*

In this direction, another formal argument is that the Wasserstein space has nonnegative curvature if the underlying space M has nonnegative curvature, but the space of maps in the Euclidean space is flat. Therefore, the two Euler-Lagrange equations (2.2) lead to a different evolution equations: for instance, if M is the Euclidean space then the Euler-Lagrange equation for the second model is simply $\ddot{\varphi} = 0$, which is a priori different from the splines Euler-Lagrange equation in the Wasserstein case.

Since, as is well-known in standard optimal transport, the Monge formulation is not well-posed, we propose to solve instead a relaxation of the problem which takes the form:

Definition 2 (Kantorovich relaxation). Let $0 = t_1 < \dots < t_n = 1$, $n \geq 3$ and $\rho_1, \dots, \rho_n \in \mathcal{P}(M)$ be n probability measures on M .

Minimize on the space of probability measures on the path space $H^2([0, 1], M)$ denoted by \mathcal{P} in short,

$$(4.4) \quad \min_{\mu} \int_{\mathcal{P}} |\ddot{x}|^2 d\mu(x),$$

which is a linear functional of $d\mu$. The curves of densities is given by its marginals in time

$$(4.5) \quad t \mapsto \rho(t)\mu_0 := [e_t]_*(\mu)$$

The marginal constraint at given time t_i are

$$(4.6) \quad [e_{t_i}]_*(\mu) = \rho_i \mu_0.$$

By standard arguments, the Kantorovich relaxation admits minimizers under general hypothesis on the manifold M , which we do not detail here. It is straightforward to check that existence of minimizers holds when $M = \mathbb{R}^d$.

As expected, the Kantorovich formulation is the relaxation of the Monge formulation, which we prove below.

Theorem 1. *Let $M = \mathbb{R}^d$, $0 = t_1 < \dots < t_n = 1$, $n \geq 3$ and $\rho_1, \dots, \rho_n \in \mathcal{P}(M)$ be n atomless probability measures on \mathbb{R}^d with compact supports. Then, under the constraints (4.6), the infimums of the variational problem (4.1) and (4.4) coincide, moreover, the infimum is attained for the latter.*

Proof. See the proof in Appendix A. □

For the proof of the theorem, we will need the following lemmas and corollaries, whose proofs are straightforward.

Lemma 2. *When the supports of the measures ρ_i are compact on \mathbb{R}^d , every minimizing μ in Definition 2 charges only the piecewise cubic polynomial functions completely by the data $p(t_i) = x_i$ for $(x_1, \dots, x_n) \in \text{Supp}(\rho_1) \times \dots \times \text{Supp}(\rho_n)$.*

Proof. The constraints are marginal constraints $[e_{t_i}]_*(\mu) = \rho_i$ for $i \geq 3$ which implies that set of paths charged by an optimal measures satisfies $x(t_i) \in \text{Supp}(\rho_i)$. In particular, any path in this set can be replaced by its minimal spline energy (which we call *cubic interpolant*) completely determined by its evaluations at times t_i for $1 \leq i \leq n$ which is unique in Euclidean space and we will denote it by c_{x_1, \dots, x_n} in the next proof. \square

Corollary 3. *As a consequence, the set of paths charged by an optimal map are uniformly C^2 and for every smooth function η with compact support, the map $t \mapsto \langle \mu(t), \eta \rangle$ is C^2 .*

Proof. The set of cubic interpolants is compact since the map $(x_1, \dots, x_n) \mapsto c_{x_1, \dots, x_n}$ is continuous from \mathbb{R}^{dn} to the space of C^2 functions (solution of an invertible linear system) and $\text{Supp}(\rho_i)$ are compact. Therefore, the set of maps are uniformly C^1 . The last point follows directly. \square

Corollary 4. *The Kantorovich problem in definition 2 on \mathbb{R}^d reduces to a multimarginal optimal transport problem, as follows, let $c(x_1, \dots, x_n)$ be the continuous cost of the cubic interpolant at times t_1, \dots, t_n , the minimization of (4.4) reduces to the minimization of*

$$(4.7) \quad \int_{M^n} c(x_1, \dots, x_n) d\pi(x_1, \dots, x_n)$$

on the space of probability measures $\pi \in \mathcal{P}(M^d)$ and under the marginal constraints $(p_i)_*(\pi) = \rho_i$ where p_i is the projection of the i^{th} factor. We denote this problem by (K) .

Proof. Direct consequence of Lemma 2. \square

The dual formulation of the minimization problem (K) is also well known [13, Theorem 2.1]

Definition 3 (Kantorovich dual problem (KP)). Let $\mathcal{Q} = \{\varphi_i \in L^1(\rho_i \mu_0), i = 1..n\}$ be the space of integrable n -uplet. Maximize on \mathcal{Q}

$$(4.8) \quad \sum_{i=1}^n \int_M \varphi_i \rho_i \mu_0, \text{ under the constraint } \sum_{i=1}^n \varphi_i(x_i) \leq c(x_1, \dots, x_n).$$

And the following duality results holds true:

Proposition 5. *There exists a n -uplet $(\varphi_i)_{i=1..n} \in \mathcal{Q}$ optimal for (KP) . Moreover $(K) = (KP)$ and for any π optimal in (4.7) there holds $\sum_1^n \varphi_i(x_i) = c(x_1, \dots, x_n)$, π almost everywhere.*

Remark 2. *A natural question is whether the solution of the Kantorovich problem (K) is admissible in the Monge formulation (MP) (Definition 1). With the formulation reduced above the spline, given by (4.7), one can try to apply existing theory to answer to this question, see [13, 19] and references therein for precise criterion. However our cost does not satisfy any of those known criterion.*

5. DISCRETIZED PROBLEM AND NUMERICAL METHOD

In this section, we discuss numerical approaches that can be efficiently applied in this setting. At least two different methods can be proposed and are based on the discretization of the formulations in Corollary (4) and Definition (2). For computational reasons explained in the next section, we devoted most of our efforts to the latter.

5.1. Hermite interpolation. In this section, we are interested in a slightly different problem than the interpolation of a time sequence of densities. We want to specify the positions and first derivatives at time 0 and 1. That is, we consider the Kantorovich problem of minimizing (4.4) under the constraints $(\mu(0), \dot{\mu}(0))$ and $(\mu(1), \dot{\mu}(1))$ are specified. In this paragraph, we address the discretization of the problem in order to propose a numerical solution to the problem. We assume that the data are of the following type, $\mu(0) = \sum_{i=1}^k \alpha_i \delta_{x_i}$, $\dot{\mu}(0) = \sum_{i=1}^k v_i \alpha_i \delta_{x_i}$ with $v_i \in \mathbb{R}^d$ and

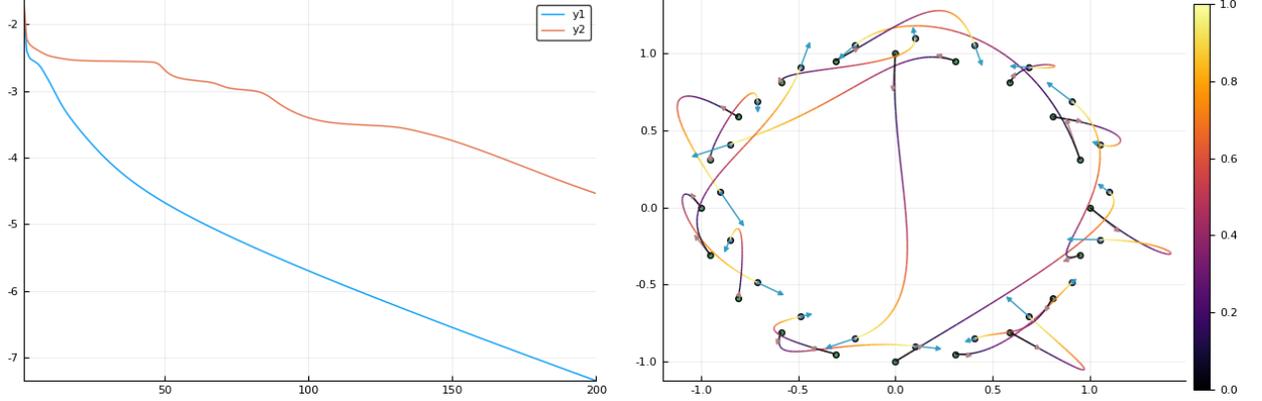


FIGURE 1. Convergence and Hermite interpolation problem.

$\mu(1) = \sum_{j=1}^l \beta_j \delta_{y_j}$, $\mu(1) = \sum_{j=1}^l \beta_j w_j \delta_{y_j}$. Then, the Kantorovich problem can be recast into a standard optimal transport problem with the following cost:

$$(5.1) \quad c((x, v), (y, w)) = 12|x - y|^2 + 4(|v|^2 + |w|^2 + \langle v, w \rangle + 3\langle v + w, x - y \rangle)$$

This cost is *not* a Riemannian cost on the tangent space of \mathbb{R}^d since if $v_x = v_y$ and x, y are close, the cost is dominated by the term $4(|v|^2 + |w|^2 + \langle v, w \rangle)$ which need not be zero. Then, the Kantorovich problem reduces to the minimization of

$$(5.2) \quad \sum_{i,j=1}^{k,l} \pi_{i,j} c((x_i, v_i), (y_j, w_j)),$$

under the constraints

$$(5.3) \quad \begin{cases} \sum_{i=1}^k \pi_{i,j} = \beta_j \\ \sum_{j=1}^l \pi_{i,j} = \alpha_i \end{cases}$$

It is straightforward to apply entropic regularization in this case, which amounts to add, for a positive parameter ε , $\varepsilon \sum_{i,j} \pi_{i,j} \log(\pi_{i,j})$ to the previous linear functional and to numerically solve the corresponding variational problem with the Sinkhorn algorithm. It is interesting to note that the choice of ε is more delicate than in the standard case of a quadratic distance cost. In Figure 1, we present the convergence rate of the method with respect to two different values of ε and the most likely deterministic plan given the optimal plan π^ε . Note that this entropic regularization method scales with the number of points as N^2 and is valid in any dimension.

In this paragraph, we discuss the extension of the entropic regularization to the cost cubic spline cost $c(x_1, \dots, x_n)$. The entropic regularization algorithm applies to any type of cost but the dimensionality of the base space cannot be greater than 4 when using a grid, for any general cost. However, when the cost is separable, such as in the multimarginal Euler problem, it is still possible to use the entropic regularization strategy. In our case, the cost $c(x_1, \dots, x_n)$ is not separable but it is possible using another parametrization to find new coordinates (z_1, \dots, z_n) and a linear invertible map A : such that the cost $c \circ A(z_1, \dots, z_n)$ is separable. Unfortunately, the constraint is changed to $[p_i]_*(A_*\pi)$ which leads to a dense adjoint operator on the Lagrange multiplier and cannot be handled efficiently.

Another possibility consists in relaxing the space of splines passing through (x_1, \dots, x_n) by allowing non optimal speed at points (x_0, \dots, x_n) , thus parametrizing by $(x_1, v_1, \dots, x_n, v_n)$. Then, in this case, the cost is:

$$(5.4) \quad c(x_0, v_0, \dots, x_n, v_n) = \sum_{i=1}^{n-1} c((x_i, v_i), (x_{i+1}, v_{i+1}))$$

where the cost c on the right-hand side is defined in Formula (5.1). The implementation of this approach requires to store a grid in (x, v) variables (say for a total number of points on the grid nk) and the associated Gibbs matrix has $(nk)^2$. This approach is very expensive in terms of memory consumption when the dimension of x is greater than 1. Moreover, the choice of the grid in the speed variable v is very sensitive and data dependent similarly to the Hermite interpolation case. This is our motivation to explore the direct discretization of Definition 2, which enables fast computation in 2D.

5.2. Discretization. In what follows, the time cylinder $[0, 1] \times M$ is discretized in time as $\bigotimes_{i=0, N} M_i$, the product space of $N + 1$ copies of M at each of the $N + 1$ time steps. We will use a regular time step discretization $\tau_i = i d\tau$ where $d\tau = \frac{1}{N}$. Using a classic finite difference approach, the time discretization of (4.4) is

$$(5.5) \quad \min_{\mu_{d\tau}} \int_{\bigotimes_{i=0, N} M_i} c_{d\tau}(x_1, \dots, x_N) d\mu_{d\tau}(x_1, \dots, x_N),$$

where $\mu_{d\tau}$ now spans the space of probability measures on $\bigotimes_{i=0, N} M_i$ representing the space of piecewise linear curves passing through x_0, x_1, \dots, x_N at times τ_0, \dots, τ_N .

A straightforward computation gives

$$(5.6) \quad c_{d\tau}(x_1, \dots, x_N) := \sum_{i=1, N-1} \frac{\|x_{i+1} + x_{i-1} - 2x_i\|^2}{d\tau^3}$$

For all times, marginals (4.5) are computed as :

$$(5.7) \quad \tau_j \mapsto \int_{\bigotimes_{i \neq j} M_i} d\mu_{d\tau}(x_1, \dots, x_N)$$

In order to simplify the presentation we will assume that the marginal constraints (4.6) are set at times t_1, \dots, t_n which coincide with times steps of the discretization (of course $n < N$, meaning the number of constraint is not the same as the number of time steps).

In short, there exist $(j_1, \dots, j_n) \in [0, N]$ such that

$$(t_1, \dots, t_n) = (\tau_{j_1}, \dots, \tau_{j_n}).$$

The constraint (4.6) becomes for all $k = 1, \dots, n$

$$(5.8) \quad \int_{\bigotimes_{i \neq j_k} M_i} d\mu_{d\tau}(x_1, \dots, x_N) = \rho_{j_k}(x_{j_k})$$

where ρ_{j_k} is the prescribed density to interpolate at time $\tau_{j_k} = t_k$.

The time discretized problem is the multimarginal problem (5.5 -5.8).

The simplest space discretization strategy is to use a regular cartesian grid. In dimension 2 and for $M = [0, 1]^2$ and at time t_i , the grid will be denoted $x_{\alpha_i, \beta_i} = (\alpha_i h, \beta_i h)$ for $(\alpha_i, \beta_i) \in [0, N_x]$ and $h = \frac{1}{N_x}$, $a = \{\alpha_i\}$ and $b = \{\beta_i\}$ will be the vectors of indices.

The time and space discretization of the problem then becomes

$$(5.9) \quad \min_T \sum_{a, b} C_{a, b} T_{a, b}$$

Where T is the $N \times N_x \times N_x$ tensor of grid values $\mu_{d\tau}(x_{\alpha_1, \beta_1}, \dots, x_{\alpha_N, \beta_N})$ and

$$(5.10) \quad C_{a, b} = c_{d\tau}(x_{\alpha_1, \beta_1}, \dots, x_{\alpha_N, \beta_N})$$

The marginals (5.7) at all times τ_j are given by

$$(5.11) \quad \sum_{a \setminus \{\alpha_j\}, b \setminus \{\beta_j\}} T_{a,b}$$

The constraints (5.8) therefore becomes for all k

$$(5.12) \quad \sum_{a \setminus \{\alpha_{j_k}\}, b \setminus \{\beta_{j_k}\}} T_{a,b} = \rho_{j_k}(x_{\alpha_{j_k}, \beta_{j_k}})$$

$a \setminus \{\alpha_{j_k}\}$ denotes the set of indices a minus α_{j_k} .

5.3. Entropic regularization and Sinkhorn algorithm. The linear programming problems (5.9-5.12) is extremely costly to solve numerically and a natural strategy, which has received a lot of attention recently following the pioneering works of [8] and [7] is to approximate these problems by strictly convex ones by adding an entropic penalization. It has been used with good results on a number of multi-marginal optimal transport problems [1] [2] [3]. Here is a rapid and simplified description, see the references above for more details.

The regularized problem is

$$(5.13) \quad \min_{T^\epsilon} \sum_{a,b} \{C_{a,b} T_{a,b}^\epsilon + \epsilon T_{a,b}^\epsilon \log(T_{a,b}^\epsilon)\}$$

It is strictly convex. Denoting $u_{\alpha_{j_k}, \beta_{j_k}}^k$ the Lagrange multipliers of the k constraints (5.12), we obtain the optimality conditions:

$$(5.14) \quad T_{a,b}^\epsilon = K_{a,b} \prod_{k=1}^N U_{j_k}^k$$

where

$$U_{j_k}^k = e^{\frac{1}{\epsilon} u_{\alpha_{j_k}, \beta_{j_k}}^k} \quad K_{a,b} = e^{-\frac{1}{\epsilon} C_{a,b}}$$

Equation (5.14) characterize the optimal tensor as a scaling of the Kernel K depending on the dual unknown U^k . Inserting this factorization into the constrains (5.12) the dual problem takes the form of the set of equations ($\forall k \in [1, n]$)

$$(5.15) \quad U_{j_k}^k = \rho_{j_k}(x_{\alpha_{j_k}, \beta_{j_k}}) \left(\sum_{a \setminus \{\alpha_{j_k}\}, b \setminus \{\beta_{j_k}\}} K_{a,b} \prod_{k' \in \{1, \dots, n\} \setminus k} U_{j_{k'}}^{k'} \right)^{-1}$$

Sinkhorn algorithm simply amounts to perform a Gauss-Seidel type iterative resolution of the system (5.15) and therefore consists in computing the sums on the right-hand side and then perform the (grid) point wise division.

5.4. Implementation. In dimension 2, each unknown U_k has dimension N_x^2 , the cost of one full Gauss Seidel cycle, i.e. on Sinkhorn iteration on all unknowns, will therefore be $n \times N_x^2 \times$ the cost to compute the tensor matrix products in the denominator of (5.15). Remember that n is the number of time steps with constraints and N the total number of time steps. The given tensor Kernel $K_{a,b}$ is a priori a large $N \times N_x \times N_x$ tensor with indices $a, b = \alpha_1, \dots, \alpha_N, \beta_1, \dots, \beta_N$. It can however advantageously be tensorized both along dimensions and also margins. First, using (5.6-5.10) we see that the Kernel is the product of smaller tensors

$$K_{a,b} = \prod_{i=1, N-1} K_{i-1, i, i+1}^0, \text{ with } K_{i-1, i, i+1}^0 := e^{-\frac{1}{\epsilon d \tau^3} \|x_{\alpha_{i+1}, \beta_{i+1}} + x_{\alpha_{i-1}, \beta_{i-1}} - 2x_{\alpha_i, \beta_i}\|^2}.$$

Moreover as we chose to work on a cartesian grid at all time steps, K^0 tensorize again into

$$K_{i-1, i, i+1}^0 = K_{i-1, i, i+1}^\alpha K_{i-1, i, i+1}^\beta \text{ with } K_{i-1, i, i+1}^\alpha := e^{-\frac{h^2}{\epsilon d \tau^3} \|\alpha_{i+1} + \alpha_{i-1} - 2\alpha_i\|^2}$$

Finally our large kernel $K_{a,b}$ can be represented a the product of $2(N-2)$ identical tensors of size $N_x \times N_x \times N_x$. Assuming a cubic cost n^3 for the multiplication of two $(n \times n)$ matrix, we see oru

algorithm is of order $O(N N_x^4)$ in dimension 2. A similar analysis

Remark 3. *A small improvement in dimensionnality can be obtained by replacing the cost (5.6) by*

$$(5.16) \quad c_{d\tau}(x_1, \dots, x_N) := \frac{1}{d\tau^3} \sum_{i=1, N-1} 2 \|x_{i+1} - x_i\|^2 + 2 \|x_i - x_{i-1}\|^2 - \|x_{i+1} - 2x_{i-1}\|^2$$

which yields the same minimization problem.

5.5. Numerical Simulations.

1D case: We present a 1D test case to highlight some of the qualitative properties of the cubic splines interpolation on the space of densities.

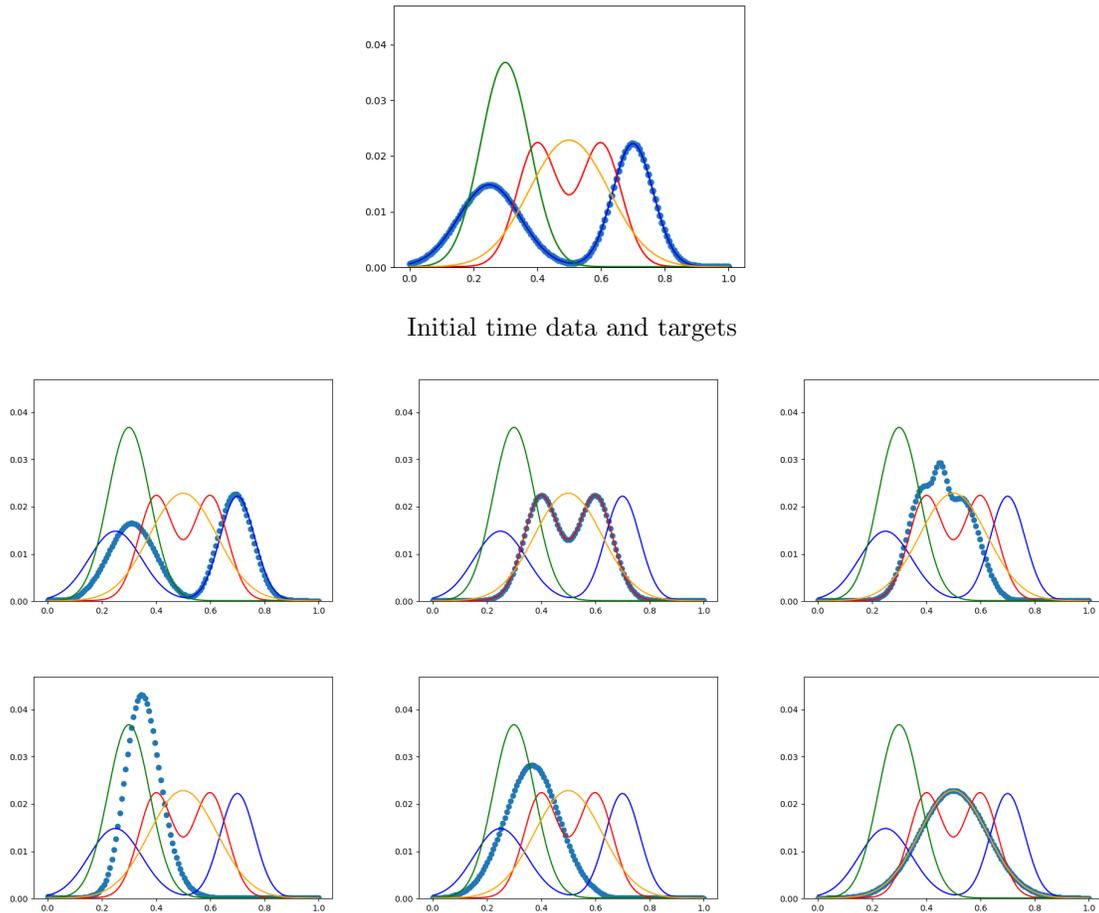


FIGURE 2. Four interpolation timepoints, 1, 6, 11, 16 and representation of the four density configurations, as well as 6 intermediate times. This experiment underlines that the spline curve has more smoothness in time and can present some concentration or diffusion effects depending on the data. The entropic regularization parameter is $\varepsilon = 8.10^{-5}$.

We consider four interpolation time points and the corresponding data are mixture of Gaussians of different standard deviations. We use a discretization of 140 points on the interval $[0, 1]$ with 16 time steps. This experiment shows that the mass can concentrate or diffuse in some situation.

Another important point here is that the entropic regularization parameter has an important impact on this concentration/diffusion effects: we show the simulations for $\varepsilon = 0.002$ and $\varepsilon = 8.10^{-5}$. In the simulation with a large ε , the concentration effect is not present and it is due to the diffusion on the path space

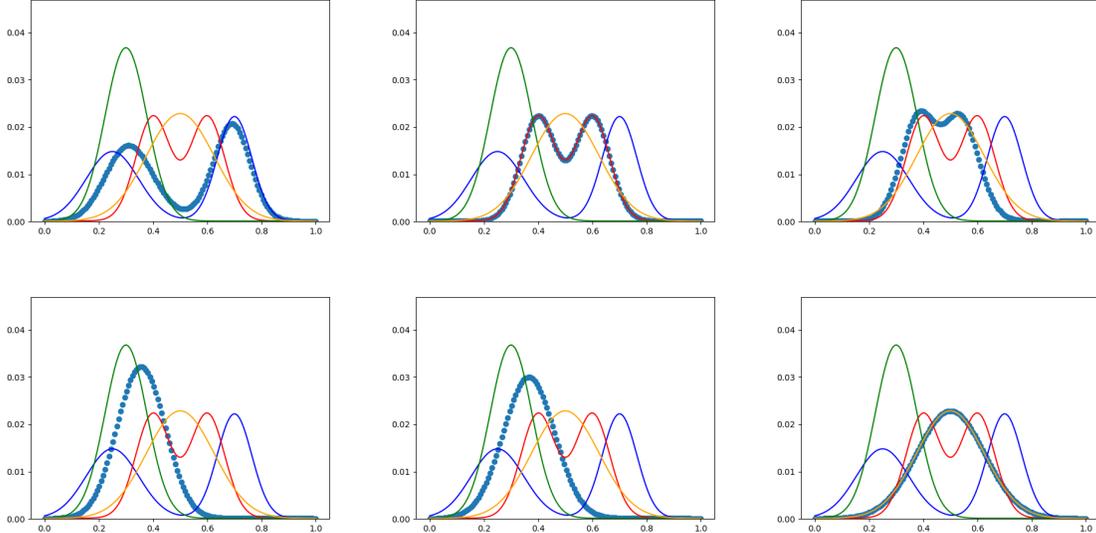


FIGURE 3. The same experiment with a larger entropic regularization parameter $\varepsilon = 0.002$. As expected, we observe less concentration of mass.

2D case: We present a 2D test case which computes a Wasserstein spline in the sense of (5.9) interpolating four Gaussian identical densities at time 1, 5, 13, and 17, see figure 4. We use a time step $d\tau = 1$ and 17 $N = 17$ time steps. The space discretization is $Nx = 50$. The entropic regularization parameter is $\varepsilon = 0.002$, note that the stability of the method depends on this parameter. It also generates artificial diffusion as it becomes more costly to concentrate the available mass on fewer Euclidean splines between the points of the support of the four Gaussians. We can compute the interpolating densities at intermediate times using (5.11) but it is more interesting to represent in figure 5 the contour line of the third quartile, i.e. the highest values of the densities representing 1/4 of the total mass. It is reasonable to say for smaller Entropic regularization parameters that we cannot reach with the current space discretization the mass will concentrate in this region.

We compare this solution with the classical Quadratic cost Optimal Transport interpolation, i.e. with the speed instead of the acceleration in the cost. More precisely taking :

$$(5.17) \quad c_{d\tau}(x_1, \dots, x_N) := \sum_{i=0, N-1} \frac{\|x_{i+1} - x_i\|^2}{d\tau}$$

As expected the mass follows respectively the linear interpolation or the Euclidean spline interpolation of the center of the Gaussians which are represented as thick red lines in figure 4.

Finally we show the convergence of the Sinkhorn iterate for both simulations in figure 5. The convergence is much slower for the speed case but we did not optimize the implementation which does not need tensors and instead just used a degraded version of the acceleration code. This may be the reason for this strange difference.

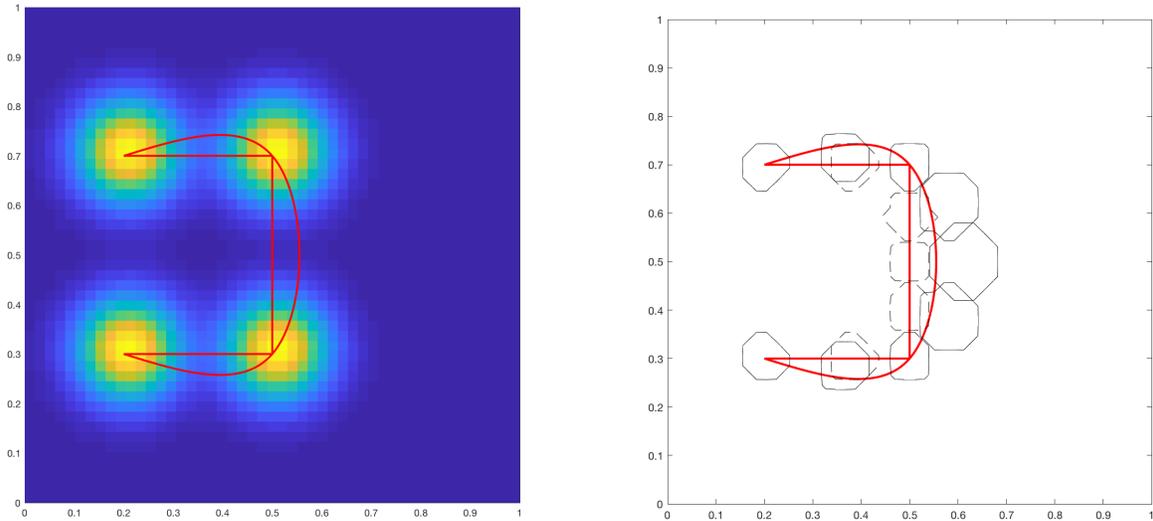


FIGURE 4. Spline interpolation of Four Gaussians with 17 times steps. Left : the data and the linear and classic cubic spline interpolation of the of Gaussian center point. Right : the level curve of the third quartile of the density every 2 time steps, in solid line for our Spline Wasserstein interpolation and in dashed line for the classic quadratic cost (speed) interpolation.

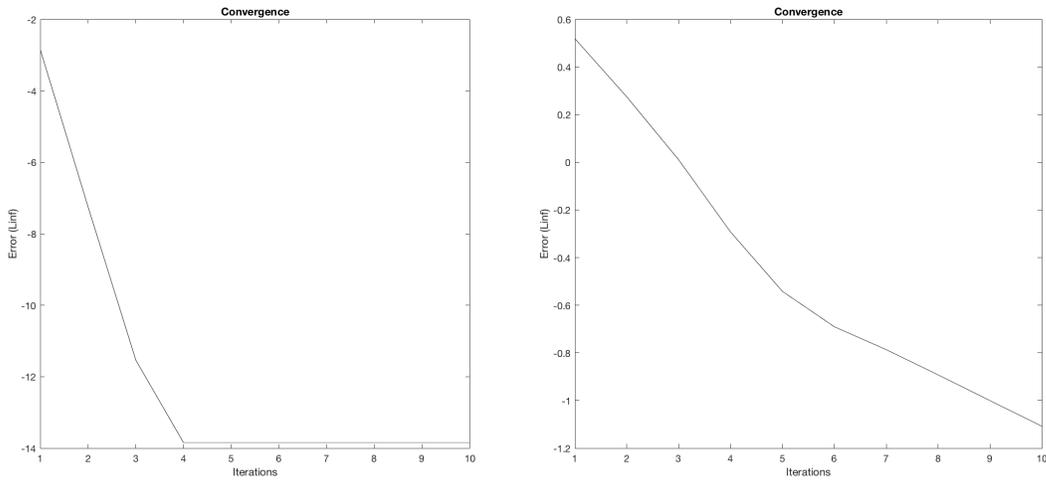


FIGURE 5. Convergence, i.e. Infinity norm of the difference of the Dual unknown between to Sinkhorn iteration. This is computed every 10 iterations. Left :for the acceleration cost, right : for the speed cost .

5.6. Extrapolation. The minimization of the acceleration can be used to provide time extension of Wasserstein geodesic in a natural way: particles follow straight lines. This can be implemented in a 3-marginal problem with the acceleration cost $c(x_1, x_2, x_3) = |x_3 - 2x_2 + x_1|^2$ under marginal constraints at time 1 and 2. Note that, in the spline model, the formulation we proposed does

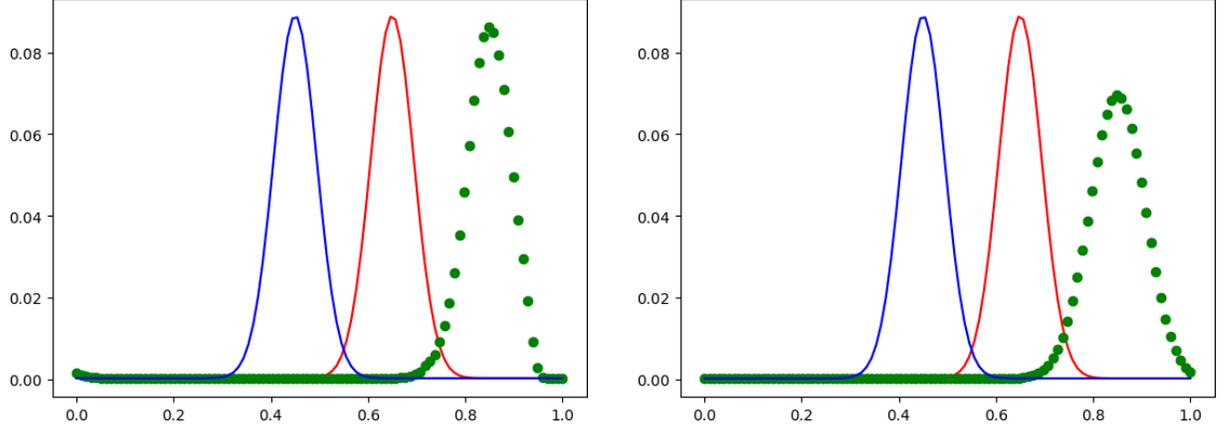
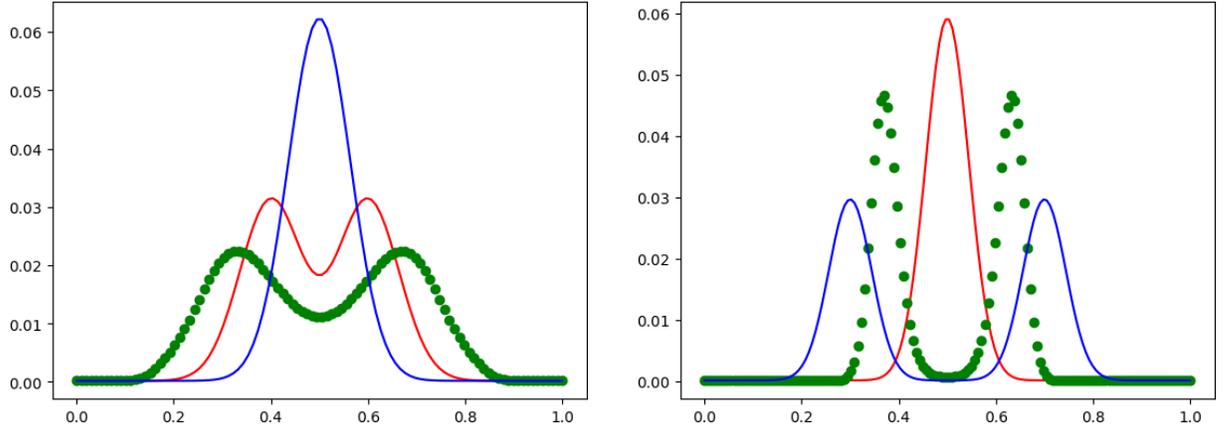
FIGURE 6. Extrapolation of a translation with two different $\varepsilon = 0.015$ and $\varepsilon = 0.03$ 

FIGURE 7. On the left, a splitting experiment and on the right, a merging experiment.

not prevent particles from crossing each other. They are completely independent. Therefore, the particles following simply geodesic lines and after a shock, the evolution is not geodesic in the Wasserstein sense (since shocks do not occur but at initial and final times). The implementation of time extrapolation using entropic regularization is straightforward. Figures 6 and 7 show some experiments on $[0, 1]$ discretized with 100 points and $\varepsilon = 0.015$. The translation experiment recovers what is expected however the effect of the diffusion can be seen with a twice larger ε . We also show two other simulations, one is a splitting simulation and the last one is a merging of two "bumps" into a single one. The extrapolation shows an other bimodal distribution which is explained by particle crossings. Note that this extrapolation scheme may proven useful in the development of higher-order scheme for the JKO algorithm.

6. SEMI-DISCRETE APPROACH

We propose another numerical scheme based on the semi-discrete approach introduced by Mérigot in [16] in dimension 2 and developed by Levy [14] in dimension 3. Here we approximate the optimal plan π in the formulation (4.7) by a sum of N tensor product of diracs masses, That is $\pi_N = \sum_{j=1}^N \left(\bigotimes_{i=1}^n \frac{1}{N} \delta_{X_j^i} \right) = \sum_{j=1}^N \frac{1}{N} \delta_{(X_j^1, \dots, X_j^n)}$. We then have to relax the constraint $(p_i)_*(\pi) = \rho_i$

since $(p_i)_*(\pi_N) = \sum_{j=1}^N \frac{1}{N} \delta_{X_j^i}$ cannot be absolutely continuous. It leads to the following variational problem.

Definition 4 (Semi-discrete variational problem). Let $\epsilon > 0$, $0 = t_1 < \dots < t_n = 1$, $n \geq 3$ and $(\rho_i)_{i=1\dots n}$ be n absolutely continuous measure. Recall that $c(Y_1, \dots, Y_n)$ is the norm of the cubic spline passing through the points (Y_1, \dots, Y_n) at time (t_1, \dots, t_n) . Let

$$\mathcal{Q}^N = \left\{ \sum_{j=1}^N \frac{1}{N} \delta_{(X_j^1, \dots, X_j^n)} \mid (X_j)_{j=1, \dots, N} \in M^n \right\}.$$

Then the semi-discrete variational problem, (SDV), is given by

$$(6.1) \quad (SDV) = \min_{\mathcal{Q}^N} \sum_{j=1}^N c(X_j^1, \dots, X_j^n) + \sum_{i=1}^n \frac{1}{\epsilon^2} W_2^2 \left(\sum_{j=1}^N \frac{1}{N} \delta_{X_j^i}, \rho_i \right),$$

where W_2 is the classical Wasserstein distance given by the quadratic cost.

Numerically the computation of W_2 is done with Merigot in dimension 2 and Levy in dimension 3 solvers. The computation of A is given via a scilab function. The gradient is computed with automatic differentiation.

Remark 4. *The advantages of the semi-discrete implementation are that one can work in dimension 2 and 3 and also consider higher order spline regularity (just replace A by the L^2 norm of the derivative of order m). One disadvantage is that the problem (SDV) is not convex.*

7. PERSPECTIVES

In this paper, we presented natural approaches to define cubic splines on the space of probability measures. We have presented a Monge formulation and its Kantorovich relaxation on the path space as well as their corresponding reduction on minimal cubic spline interpolation. We leave for future work theoretical questions such as the study of conditions under which the existence of a Monge map as a minimizer occurs, as well as the relaxation of cubic spline in the Wasserstein metric. Our main contributions focus on the numerical feasibility of the minimization of the acceleration on the path space with marginal constraints. We have developed the entropic regularization scheme for the acceleration and shown simulations in 1D and 2D. Future work will address the 3D case which is out of reach with the methods presented in the first sections of this paper but possibly tackled with the semi-discrete method presented in Section 6. In a similar direction, the application of this approach to the unbalanced case in the spirit of [5] seems challenging due to the this dimensionality constraint and could be achieved within the semi-discrete setting.

APPENDIX A. PROOF OF THEOREM 1

Proof. We first reduce the dimensionality of the two problems in both formulations by Corollary 4 on the space of cubic interpolants. We now adapt the proof of [20, Thm 1.5.5] to prove the result. Consider an optimal measure μ on the path space. By lemma 2, it only charges the space of cubics interpolants c_{x_1, \dots, x_n} . Without loss of generality, we can consider that the points x_i belong to a common compact set K , on which we choose an arbitrary measurable partition $(K_s^\epsilon)_{s \in S}$ such that $\text{diam}(K_s) \leq \epsilon$. We denote by $\mathcal{K}_{s_1, \dots, s_n}$ an element of the n -product of the partition.

We use an induction argument, the problem is trivial when $n = 2$ and the existence of a map approximating a coupling plan follows from [20, Thm 1.5.5]. Our induction hypothesis is the existence of a plan π^ϵ on M^n such that it is deterministic and it approximates the cost of the original plan π at ϵ . A plan π is deterministic if it can be written as $\gamma_T = \delta_{x_1, T_1 x_1, \dots, T_n x_1} \rho_1(x_1)$, where T_1, \dots, T_n are measurable maps on M . We now show the hypothesis is true for $n + 1$ by conditioning at time t_1 , that is we can restrict the plan π on $K_{s_1}^\epsilon \times K^n$ denoted by $\gamma_i = \pi_{K_{s_1}^\epsilon \times K^n}$ and find a map T_s^ϵ that pushforwards $(e_1)_*(\gamma_i)$ onto $(e_2)_*(\gamma_i)$ since the two marginals are atomless. Moreover, for each plan $(e_{2, \dots, n+1})_*(\gamma_i)$ (which is atomless), by induction there exist maps $[T_{s_i}^\epsilon]$ as above

which depends on γ_i . Then, T_1^ε can be defined as the concatenation of the maps $T_{s_i}^\varepsilon$ since the sets K_{s_i} are disjoint, as well as the other maps T_k^ε . By construction, the plan denoted by γ_T^ε satisfies $|\langle \gamma_T^\varepsilon(\mathcal{K}_{s_1, \dots, s_n}) - \pi(\mathcal{K}_{s_1, \dots, s_n}), c \rangle| \leq \varepsilon$ by direct check as in [20, Thm 1.5.5].

This denseness property is sufficient to prove Theorem 1 since the approximating map $\varphi(t)$ is the evaluation at time t of the deterministic plan $\delta_{x_1, T_1^\varepsilon x_1, \dots, T_n^\varepsilon x_1}$. \square

REFERENCES

- [1] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- [2] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. A Numerical Method to solve Optimal Transport Problems with Coulomb Cost. working paper or preprint, May 2015.
- [3] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm. working paper or preprint, October 2017.
- [4] M. Camarinha, F. Silva Leite, and P.Crouch. Splines of class C^k on non-euclidean spaces. *IMA Journal of Mathematical Control & Information*, 12:399–410, 1995.
- [5] L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard. An Interpolating Distance between Optimal Transport and Fisher-Rao. *Found. Comp. Math.*, 2016.
- [6] P. Crouch and F. Silva Leite. The dynamic interpolation problem: On Riemannian manifold, Lie groups and symmetric spaces. *Journal of dynamical & Control Systems*, 1:177–202, 1995.
- [7] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.
- [8] Alfred Galichon and Bernard Salanié. Matching with Trade-offs: Revealed Preferences over Competing Characteristics. working paper or preprint, April 2010.
- [9] F. Gay-Balmaz, D. D. Holm, D. M. Meier, T. S. Ratiu, and F.-X. Vialard. Invariant Higher-Order Variational Problems. *Communications in Mathematical Physics*, 309:413–458, January 2012.
- [10] F. Gay-Balmaz, D. D. Holm, D. M. Meier, T. S. Ratiu, and F.-X. Vialard. Invariant Higher-Order Variational Problems II. *Journal of NonLinear Science*, 22:553–597, August 2012.
- [11] B. Heeren, M. Rumpf, and B. Wirth. Variational time discretization of Riemannian splines. *ArXiv e-prints*, November 2017.
- [12] B. Khesin and R. Wendt. *The geometry of infinite-dimensional groups*, volume 51. Springer Science & Business Media, 2008.
- [13] Young-Heon Kim and Brendan Pass. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.
- [14] Lévy, Bruno. A numerical algorithm for l2 semi-discrete optimal transport in 3d. *ESAIM: M2AN*, 49(6):1693–1715, 2015.
- [15] J. Lott. Some geometric calculations on Wasserstein space. *Communications in Mathematical Physics*, 277(2):423–437, 2008.
- [16] Quentin Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30 (5):1583–1592, 2011.
- [17] L. Noakes, G. Heinzinger, and B. Paden. Cubic splines on curved spaces. *IMA Journal of Mathematical Control & Information*, 6:465–473, 1989.
- [18] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.
- [19] Pass, Brendan. Multi-marginal optimal transport: Theory and applications. *ESAIM: M2AN*, 49(6):1771–1790, 2015.
- [20] F. Santambrogio. Optimal transport for applied mathematicians. *Progress in Nonlinear Differential Equations and their applications*, 87, 2015.
- [21] Nikhil Singh, François-Xavier Vialard, and Marc Niethammer. Splines for diffeomorphisms. *Medical Image Analysis*, 25(1):56–71, 2015.
- [22] R. Tahraoui and F.-X. Vialard. Riemannian cubics on the group of diffeomorphisms and the Fisher-Rao metric. *ArXiv e-prints*, June 2016.
- [23] F.-X. Vialard and A. Trounev. Shape Splines and Stochastic Shape Evolutions: A Second Order Point of View. *Quart. Appl. Math.*, 2012.

INRIA, PROJECT TEAM MOKAPLAN
E-mail address: `jean-david.benamou@inria.fr`

INRIA, PROJECT TEAM MOKAPLAN
E-mail address: `thomas.gallouet@inria.fr`

UNIVERSITÉ PARIS-DAUPHINE, PSL RESEARCH UNIVERSITY, CEREMADE, INRIA, PROJECT TEAM MOKAPLAN
E-mail address: `fxvialard@normalesup.org`