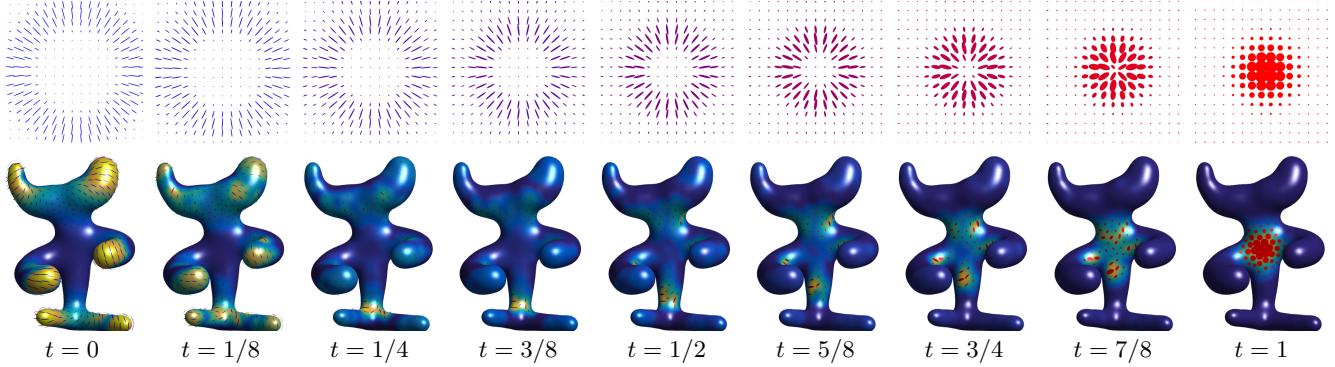


# Quantum Optimal Transport for Tensor Field Processing



**Figure 1:** Given two input fields of positive semidefinite matrices (displayed at times  $t \in \{0, 1\}$  using ellipses) on some domain (here, a 2-D planar square and a surface mesh), our Quantum Optimal Transport (Q-OT) method defines a continuous interpolating path for  $t \in [0, 1]$ . Unlike linear interpolation schemes, Q-OT transports the “mass” of the tensors (size of the ellipses) as well as their anisotropy and orientation. This interpolation, and its extension to finding the barycenter of several input fields, is computed using a fast extension of the well-known Sinkhorn algorithm.

## 1 Abstract

This article introduces a new notion of optimal transport (OT) between tensor fields, which are measures whose values are positive semidefinite (PSD) matrices. This “quantum” formulation of OT (Q-OT) corresponds to a relaxed version of the classical Kantorovich transport problem, where the fidelity between the input PSD-valued measures is captured using the geometry of the Von-Neumann quantum entropy. We propose a quantum-entropic regularization of the resulting convex optimization problem, which can be solved efficiently using an iterative scaling algorithm. This method is a generalization of the celebrated Sinkhorn algorithm to the quantum setting of PSD matrices. We extend this formulation and the quantum Sinkhorn algorithm to compute barycenters within a collection of input tensor fields. We illustrate the usefulness of the proposed approach on applications to procedural noise generation, anisotropic meshing, diffusion tensor imaging and spectral texture synthesis.

**Keywords:** Optimal transport, tensor field, PSD matrices, quantum entropy

**Concepts:** •Computing methodologies → Shape analysis;

## 1 Introduction

Optimal transport (OT) is an active field of research at the intersection of probability theory, PDEs, convex optimization and numerical analysis. OT offers a canonical way to lift a ground distance on some metric space to a metric between arbitrary probability measures defined over this base space. OT distances offer many interesting features, and in particular lead to a geometrically faithful way to manipulate and interpolate probability distributions.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). © 2017 Copyright held by the owner/author(s).

### 28 1.1 Previous Work

**29 Scalar-valued optimal transport.** Dating back to the eighteenth century, classical instances of the optimal transport problem seek 30 a minimal-cost matching between two distributions defined over 31 a geometric domain, e.g. matching supply to demand while incurring 32 minimal cost. Initially formulated by Monge in terms of an 33 unknown map transporting mass [1781], its reformulation by Kan- 34 torovich [1942] as a linear program (static formulation) enables the 35 use of convex analysis to study its structure and develop numerical 36 solvers. The equivalence between these two formulations was 37 introduced by Brenier [1991] and opened the door to a dynamical 38 (geodesic) reformulation [Benamou and Brenier 2000]. We refer 39 to [Santambrogio 2015] for a review of the theoretical foundations 40 of OT.

The basic OT problem has been extended in various ways, a typical 42 illustration of which being the computation of a barycenter (Fréchet 43 mean) of input measures, a convex program studied by Aguech and 44 Carlier [2011]. OT has found numerous applications, for instance in 45 computer vision (under the name “earth mover distance”) [Rubner 46 et al. 2000] or computer graphics [Bonneel et al. 2011].

**48 Unbalanced transport.** While the initial formulations of OT are 49 restricted to positive measures of equal mass (normalized probability 50 distributions), a recent wave of activity has proposed and studied 51 a family of “canonical” extensions to the “unbalanced” setting of 52 arbitrary positive measures. This covers both a dynamic formulation 53 [Liero et al. 2016; Kondratyev et al. 2015; Chizat et al. 2016b] 54 and a static one [Liero et al. 2015; Chizat et al. 2015] and has been 55 applied in machine learning [Frogner et al. 2015]. Our work extends 56 this static unbalanced formulation to tensor-valued measures.

**57 Entropic regularization.** The current state-of-the-art OT approx- 58 imation for arbitrary ground costs uses entropic regularization of 59 the transport plan. This leads to strictly convex programs that can 60 be solved using a simple class of highly parallelizable “diagonal 61 scaling” algorithms. The landmark paper of Cuturi [2013] inspired 62 detailed study of these solvers, leading to various generalizations 63 of Sinkhorn’s algorithm [1964]. This includes for instance the use

fast convolutional structures [Solomon et al. 2015], extensions to barycenters [Benamou et al. 2015] and to unbalanced OT [Frogner et al. 2015; Chizat et al. 2016a]. These entropic regularization techniques correspond to the use of projection and proximal maps for the Kullback–Leibler Bregman divergence and are equivalent to iterative projections [Bregman 1967] and Dykstra’s algorithm [Dykstra 1983; Bauschke and Lewis 2000]. An important contribution of the present work is to extend these techniques to the matrix setting (i.e., using quantum divergences). Note that quantum divergences have been recently used to solve some machine learning problems [Dhillon and Tropp 2008; Kulis et al. 2009; Chandrasekaran and Shah 2016].

**Tensor field processing.** Tensor-valued data are ubiquitous in various areas of imaging science, computer graphics and vision. In medical imaging, diffusion tensor imaging (DTI) directly maps observed data to fields of tensors [Dryden et al. 2009; Deriche et al. 2006]. Tensor fields are also at the heart of anisotropic diffusions techniques in image processing [Weickert 1998] and find applications to line drawing [Vaxman et al. 2016] and data visualization [Hotz et al. 2004].

**OT on tensor fields.** The simplest way to define OT-like distances between arbitrary vector-valued measures is to use dual norms [Ning and Georgiou 2014], which correspond to generalizations of  $W_1$  OT for which transport cost equals ground distance. The corresponding metrics, however, have degenerate behavior in interpolation and barycenter problems (much like the  $L^1$  norm on functions) and only use the linear structure of matrices rather than their multiplicative structure. More satisfying notions of OT have recently been proposed in a dynamical (geodesic) way [Jiang et al. 2012; Carlen and Maas 2014; Chen et al. 2016]. A static formulation of a tensor-valued OT is proposed in [Ning et al. 2015], but it differs significantly from ours. It is initially motivated using a lifting that squares the number of variables, but a particular choice of cost reduces the computation to the optimization of a pair of couplings. In contrast, the formulation we propose in the present article is a direct generalization of unbalanced OT to matrices, which in turn enables the use of a Sinkhorn algorithm.

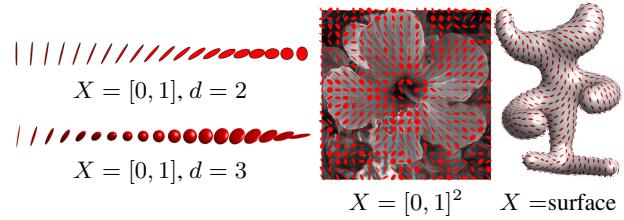
## 1.2 Contributions

We present a new static formulation of OT between tensor fields, which is the direct generalization of unbalanced OT from the scalar to the matrix case. Our second contribution is a fast entropic scaling algorithm generalizing the celebrated Sinkhorn iterative scheme. This leads to a method to compute geometrically-faithful interpolations between two tensor fields. Our third contribution is the extension of this approach to compute barycenters between several tensor fields. The Matlab code to reproduce the results of this article is available online.<sup>1</sup>

## 1.3 Notation

In the following, we denote  $\mathcal{S}^d \subset \mathbb{R}^{d \times d}$  the space of symmetric matrices,  $\mathcal{S}_+^d$  the closed convex cone of positive semidefinite matrices, and  $\mathcal{S}_{++}^d$  the open cone of positive definite matrices. We denote  $\exp : \mathcal{S}^d \rightarrow \mathcal{S}_{++}^d$  the matrix exponential, which is defined as  $\exp(P) = U \text{diag}_s(e^{\sigma_s})U^\top$  where  $P = U \text{diag}_s(\sigma_s)U^\top$  is an eigendecomposition of  $P$ . We denote  $\log : \mathcal{S}_{++}^d \rightarrow \mathcal{S}^d$  the matrix logarithm  $\log(P) = U \text{diag}_s(\log \sigma_s)U^\top$ , which is the inverse of  $\exp$  on  $\mathcal{S}_{++}^d$ .

<sup>1</sup>Removed for paper submission.



**Figure 2:** Displays of various types of tensor-valued measures  $\mu$ . The principal directions of an ellipse at some  $x_i \in X$  are the eigenvectors of  $\mu_i \in \mathcal{S}_+^d$ , while the principal width are given by its eigenvalues.

A tensor-valued measure  $\mu$  defined on some space  $X$  is a vector-valued measure, where the “mass”  $\mu(A) \in \mathcal{S}_+^d$  associated to a measurable set  $A \subset X$  is a PSD matrix. In this article, in order to derive computational schemes, we focus on discrete measures. Such a measure  $\mu$  is a sum of Dirac masses  $\mu = \sum_{i \in I} \mu_i \delta_{x_i}$ , where  $(x_i)_i \subset X$ , and  $(\mu_i)_i \in \mathcal{S}_+^d$  is a collection of PSD matrices. In this case,  $\mu(A) = \sum_{x_i \in A} \mu_i$ . Figure 2 shows graphically some examples of tensor-valued measures; we use this type of visualization through the article. In the following, since the sampling points  $(x_i)_i$  are assumed to be fixed and clear from the context, to ease readability, we do not make the distinction between the measure  $\mu$  and the vector of matrices  $(\mu_i)_i$ . This is an abuse of notation, but it is always clear from context whether we are referring to a measure or a vector.

The quantum entropy (also called von Neumann entropy) of a tensor-valued measure is

$$H(\mu) \stackrel{\text{def.}}{=} \sum_i H(\mu_i) \quad \text{where} \quad (1)$$

$$\forall P \in \mathcal{S}^d, \quad H(P) \stackrel{\text{def.}}{=} -\text{tr}(P \log(P) - P) - \iota_{\mathcal{S}_{++}^d}(P).$$

Note that  $H$  is a concave function. The quantum Kullback–Leibler divergence (also called quantum relative entropy) is the Bregman divergence associated to  $-H$ . For a collection of PSD matrices  $\mu = (\mu_i)_i, \xi = (\xi_i)_i$  in  $\mathcal{S}_+^d$  corresponding to measures defined on the same grid, assuming  $\xi_i \succ 0$ , it is defined as

$$\text{KL}(\mu|\xi) \stackrel{\text{def.}}{=} \sum_i \text{KL}(\mu_i|\xi_i), \quad (2)$$

where for all  $(P, Q) \in \mathcal{S}_+^d \times \mathcal{S}_{++}^d$ , we denote

$$\text{KL}(P|Q) \stackrel{\text{def.}}{=} \text{tr}(P(\log(P) - \log(Q)) - P + Q) + \iota_{\mathcal{S}_{++}^d}(P)$$

which is convex with respect to both arguments. The inner product between collections of matrices  $\mu = (\mu_i)_i, \xi = (\xi_i)_i$  is

$$\langle \mu, \xi \rangle \stackrel{\text{def.}}{=} \sum_i \langle \mu_i, \xi_i \rangle \stackrel{\text{def.}}{=} \sum_i \text{tr}(\mu_i \xi_i^\top).$$

Given a collection of matrices  $\gamma = (\gamma_{i,j})_{i \in I, j \in J}$  the marginalization operators read

$$\gamma \mathbb{1}_J \stackrel{\text{def.}}{=} \left( \sum_j \gamma_{i,j} \right)_i \quad \text{and} \quad \gamma^\top \mathbb{1}_I \stackrel{\text{def.}}{=} \left( \sum_i \gamma_{i,j} \right)_j.$$

## 2 Kantorovich Problem for Tensor-Valued Transport

We consider two measures that are sums of Dirac masses

$$\mu = \sum_{i \in I} \mu_i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{j \in J} \nu_j \delta_{y_j} \quad (3)$$

where  $(x_i)_i \subset X$  and  $(y_j)_j \subset Y$ , and  $(\mu_i)_i \in \mathcal{S}_+^d$  and  $(\nu_j)_j \in \mathcal{S}_+^d$  are collections of PSD matrices. Our goal is to propose a new definition of OT between  $\mu$  and  $\nu$ .

### 2.1 Tensor Transportation

Following the initial static formulation of OT by Kantorovich [1942], we define a coupling  $\gamma = \sum_{i,j} \gamma_{i,j} \delta_{(x_i, y_j)}$  as a measure over the product  $X \times Y$  that encodes the transport of mass between  $\mu$  and  $\nu$ . In the matrix case,  $\gamma_{i,j} \in \mathcal{S}_+^d$  is now a PSD matrix, describing how much of mass is moved between  $\mu_i$  and  $\nu_j$ . Exact (balanced) transport would mean that the marginals  $(\gamma \mathbb{1}_J, \gamma^\top \mathbb{1}_I)$  must be equal to the input measures  $(\mu, \nu)$ . But as remarked by Ning et al. [2015], in contrast to the scalar case, in the matrix case (dimension  $d > 1$ ), in general this constraint is too strong, and there might exists no coupling satisfying these marginal constraints. We advocate in this work that the natural workaround for the matrix setting is the unbalanced case, and following [Liero et al. 2015], we propose to use a “relaxed” formulation where the discrepancy between the marginals  $(\gamma \mathbb{1}_J, \gamma^\top \mathbb{1}_I)$  and the input measures  $(\mu, \nu)$  is quantified according to some divergence between measures.

In the scalar case, the most natural divergence is the Kulback-Leibler divergence (which in particular gives rise to a natural Riemannian structure on positive measures, as defined in [Liero et al. 2016; Kondratyev et al. 2015; Chizat et al. 2016b]). We propose to make use of its quantum counterpart (2) via the following convex program

$$W(\mu, \nu) = \min_{\gamma} \langle \gamma, c \rangle + \rho_1 \text{KL}(\gamma \mathbb{1}_J | \mu) + \rho_2 \text{KL}(\gamma^\top \mathbb{1}_I | \nu) \quad (4)$$

subject to the constraint  $\forall (i, j), \gamma_{i,j} \in \mathcal{S}_+^d$ . Here  $\rho_1, \rho_2 > 0$  are constants balancing the “transport” effect versus the local modification of the matrices.

The matrix  $c_{i,j} \in \mathbb{R}^{d \times d}$  measures the cost of displacing an amount of (matrix) mass  $\gamma_{i,j}$  between  $x_i$  and  $y_j$  as  $\text{tr}(\gamma_{i,j} c_{i,j})$ . A typical cost, assuming  $X = Y$  is a metric space endowed with a distance  $d_X$ , is

$$c_{i,j} = d_X(x_i, y_j)^\alpha \text{Id}_{d \times d},$$

for some  $\alpha > 0$ . In this case, one should interpret the trace as the global mass of a tensor, and the total transportation cost is simply

$$\langle \gamma, c \rangle = \sum_{i,j} d_X(x_i, y_j)^\alpha \text{tr}(\gamma_{i,j}).$$

**Remark 1** (Classical OT). In the scalar case  $d = 1$ , (4) recovers exactly the log-entropic definition [Liero et al. 2015] of unbalanced optimal transport, which is studied numerically by Chizat et al. [2016a]. For isotropic tensors, i.e., all  $\mu_i$  and  $\nu_j$  are scalar multiples of the identity  $\text{Id}_{d \times d}$ , the computation also collapses to the scalar case (the  $\gamma_{i,j}$  are also isotropic). More generally, if all the  $(\mu_i, \nu_j)_{i,j}$  commute, they diagonalize in the same orthogonal basis, and (4) reduces to performing  $d$  independent unbalanced OT computations along each eigendirection.

**Remark 2** (Cost between single Dirac masses). When  $\mu = P \delta_x$  and  $\nu = Q \delta_x$  are two Dirac masses are the same location  $x$  and

associated to tensors  $(P, Q) \in (\mathcal{S}_+^d)^2$ , one obtains the following “metric” between tensors (assuming  $\rho_1 = \rho_2 = \rho$  for simplicity)

$$W(P \delta_x, Q \delta_x) = D(P, Q) \stackrel{\text{def.}}{=} \text{tr}(P + Q - 2\mathfrak{M}(P, Q))^{\frac{1}{2}} \quad (5)$$

where  $\mathfrak{M}(P, Q) \stackrel{\text{def.}}{=} \exp(\log(P)/2 + \log(Q)/2)$ . Unfortunately, in general  $D$  does not satisfy the triangle inequality. Note that when  $(P, Q)$  commute, one has  $D(P, Q) = \|\sqrt{P} - \sqrt{Q}\|$  which indeed satisfies the triangle inequality.

**Remark 3** (Quantum transport on curved geometries). If  $(\mu, \nu)$  are defined on a non-Euclidean space  $Y = X$ , like a smooth manifold, then formulation (4) should be handled with care, since it assumes all the tensors  $(\mu_i, \nu_j)_{i,j}$  are defined in some common basis. For smooth manifolds, the simplest workaround is to assume that these tensors are defined with respect to carefully selected orthogonal bases of the tangent planes, so that the field of bases is itself smooth. Unless the manifold is parallelizable, in particular if it has a trivial topology, it is not possible to obtain a globally smooth orthonormal basis; in general, any such field necessarily has a few singular points. In the following, we compute smoothly-varying orthogonal bases of the tangent planes following the method of Crane et al. [2010]. In this setting, the cost is usually chosen to be  $c_{i,j} = d_X(x_i, x_j)^\alpha \text{Id}_{d \times d}$  where  $d_X$  is the geodesic distance on  $X$ .

### 2.2 Quantum Transport Interpolation

Given two input measures  $(\mu, \nu)$ , we denote  $\gamma$  as a solution of (4) or, in practice, its regularized version (see (7) below). The coupling  $\gamma$  defines a (fuzzy) correspondence between the tensor fields. A typical use of this correspondence is to compute a continuous interpolation between these fields. Section 3.3 shows some numerical illustration of this interpolation. Note also that Section 4 proposes a generalization of this idea to compute an interpolation (barycenter) between more than two input fields.

Mimicking the definition of the optimal transport interpolation (the so-called McCann displacement interpolation; see for instance [Santambrogio 2015]), we propose to use  $\gamma$  to define a path  $t \in [0, 1] \mapsto \mu_t$  interpolating between  $(\mu, \nu)$ . For simplicity, we assume the cost has the form  $c_{i,j} = d_X(x_i, y_j)^\alpha \text{Id}_{d \times d}$  for some ground metric  $d_X$  on  $X = Y$ . We also suppose we can compute efficiently the interpolation between two points  $(x_i, y_j) \in X^2$  as

$$x_{i,j}^t \stackrel{\text{def.}}{=} \underset{x \in X}{\text{argmin}} (1-t)d_X^2(x_i, x) + td_X^2(y_j, x).$$

For instance, over Euclidean spaces,  $g_t$  is simply a linear interpolation, and over more general manifold, it is a geodesic segment. We also denote

$$\bar{\mu}_i \stackrel{\text{def.}}{=} \mu_i \left( \sum_j \gamma_{i,j} \right)^{-1} \quad \text{and} \quad \bar{\nu}_j \stackrel{\text{def.}}{=} \nu_j \left( \sum_i \gamma_{i,j} \right)^{-1}$$

the adjustment factors which account for the imperfect match of the marginal associated to a solution of (7); the adjusted coupling is

$$\gamma_{i,j}^t \stackrel{\text{def.}}{=} [(1-t)\bar{\mu}_i + t\bar{\nu}_j]\gamma_{i,j}.$$

Finally, the interpolating measure is then defined as

$$\forall t \in [0, 1], \quad \mu_t \stackrel{\text{def.}}{=} \sum_{i,j} \gamma_{i,j}^t \delta_{x_{i,j}^t}. \quad (6)$$

One easily verifies that this measure indeed interpolates the two input measures, i.e.  $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$ . This formula (6) generates the interpolation by creating a Dirac tensor  $\gamma_{i,j}^t \delta_{x_{i,j}^t}$  for each coupling entry  $\gamma_{i,j}$ , and this tensor travels between  $\mu_i \delta_{x_i}$  (at  $t = 0$ ) and  $\nu_j \delta_{y_j}$  (at  $t = 1$ ).

241 *Remark 4* (Computational cost). We observed numerically that, sim-  
 242 ilarly to the scalar case, the optimal coupling  $\gamma$  is sparse, meaning  
 243 that only of the order of  $O(|I|)$  non-zero terms are involved in the  
 244 interpolating measure (6). Note that the entropic regularization al-  
 245 gorithm detailed in Section 3 destroys this exact sparsity, but we  
 246 found numerically that that thresholding to zero the small entries of  
 247  $\gamma$  generates accurate approximations.

### 248 3 Quantum Sinkhorn

249 The convex program (4) defining quantum OT is computationally  
 250 challenging because it can be very large scale (problem size is  
 251  $|I| \times |\bar{J}|$ ) for imaging applications, and it involves matrix exponen-  
 252 tial and logarithm. In this section, leveraging recent advances in  
 253 computational OT initiated by Cuturi [2013], we propose to use a  
 254 similar entropy regularized strategy (see also section 1), but this time  
 255 with the quantum entropy (1).

#### 256 3.1 Entropic Regularization

257 We define an entropic regularized version of (4)

$$258 W_\varepsilon(\mu, \nu) \stackrel{\text{def.}}{=} \min_{\gamma} \langle \gamma, c \rangle + \rho_1 \text{KL}(\gamma \mathbb{1}_J | \mu) + \rho_2 \text{KL}(\gamma^\top \mathbb{1}_I | \nu) - \varepsilon H(\gamma). \quad (7)$$

258 Note that when  $\varepsilon = 0$ , one recovers the original problem (4). This  
 259 is a strongly convex program, with a unique solution. The crux of  
 260 this approach, as already known in the scalar case (see [Chizat et al.  
 261 2016a]), is that its convex dual has a particularly simple structure,  
 262 which is amenable to a simple alternating maximization strategy.

263 **Proposition 1.** *The dual problem associated to (7) reads*

$$264 W_\varepsilon(\mu, \nu) = \max_{u, v} -\text{tr} \left[ \rho_1 \sum_i (e^{u_i + \log(\mu_i)} - \mu_i) \right. \\ \left. + \rho_2 \sum_j (e^{v_j + \log(\nu_j)} - \nu_j) + \varepsilon \sum_{i,j} e^{\mathcal{K}(u, v)_{i,j}} \right], \quad (8)$$

264 where we define

$$265 \mathcal{K}(u, v)_{i,j} \stackrel{\text{def.}}{=} -\frac{c_{i,j} + \rho_1 u_i + \rho_2 v_j}{\varepsilon}. \quad (9)$$

265 Furthermore, the following primal-dual relationships hold at opti-  
 266 mality:

$$267 \forall (i, j), \quad \gamma_{i,j} = \exp(\mathcal{K}(u, v)_{i,j}). \quad (10)$$

267 *Proof.* Applying the Fenchel–Rockafellar duality theorem [Rock-  
 268 afellar 1970] to (7) leads to the dual program

$$269 \max_{u, v} -\varepsilon \text{KL}^*(\mathcal{K}_0(u, v) | \xi) - \rho_1 \text{KL}^*(u | \mu) - \rho_2 \text{KL}^*(v | \nu) - \varepsilon \text{tr}(\xi),$$

269 where here  $\text{KL}^*(\cdot | \mu)$  corresponds to the Legendre transform with  
 270 respect to the first argument of the KL divergence,  $\mathcal{K}_0(u, v)_{i,j} \stackrel{\text{def.}}{=} -\frac{\rho_1 u_i + \rho_2 v_j}{\varepsilon}$ . and  $\xi_{i,j} \stackrel{\text{def.}}{=} \exp(-c_{i,j}/\varepsilon)$  for all  $i, j$ . The following  
 272 Legendre formula leads to the desired result:

$$273 \text{KL}^*(u | \mu) = \sum_i \text{tr}(\exp(u_i + \log(\mu_i)) - \mu_i).$$

273  $\square$

#### 274 3.2 Quantum Sinkhorn Algorithm

275 It is possible to use Dykstra’s algorithm [1983] (see [Bauschke and  
 276 Lewis 2000] for its extension to Bregman divergences) to solve (8).  
 277 This corresponds to alternatively maximizing (8) with respect to  $u$

278 and  $v$ . The following proposition states that the maximization with  
 279 respect to either  $u$  or  $v$  leads to two fixed-point equations. These  
 280 fixed points are conveniently written using the log-sum-exp operator,

$$281 \text{LSE}_j(K_{i,j}) \stackrel{\text{def.}}{=} \log \sum_j \exp(K_{i,j}), \quad (11)$$

282 where the sum on  $j$  is replaced by a sum on  $i$  for  $\text{LSE}_i$ .

283 **Proposition 2.** *For  $v$  fixed (resp.  $u$  fixed), the minimizer  $u$  (resp.  $v$ )  
 284 of (8) satisfies*

$$\forall i, \quad u_i = \text{LSE}_j(\mathcal{K}(u, v)_{i,j}) - \log(\mu_i), \quad (12)$$

$$\forall j, \quad v_j = \text{LSE}_i(\mathcal{K}(u, v)_{i,j}) - \log(\nu_j), \quad (13)$$

285 where  $\mathcal{K}(u, v)$  is defined in (9).

286 *Proof.* Writing the first order condition of (8) with respect to each  
 287  $u_i$  leads to

$$288 \rho_1 e^{u_i + \log(\mu_i)} - \rho_1 \sum_j e^{\mathcal{K}(u, v)_{i,j}} = 0$$

289 which gives the desired expression. A similar expression holds for  
 290 the first order conditions with respect to  $v_j$ .  $\square$

290 A simple fixed point algorithm is then obtained by replacing in  
 291 Dykstra’s the explicit alternating minimization with respect to  $u$  and  
 292  $v$  by just one step of fixed point iterations (12) and (13). To make  
 293 the resulting fixed point contractant and ensure linear convergence,  
 294 one introduces relaxation parameters  $(\tau_1, \tau_2)$ .

295 The quantum Sinkhorn algorithm is detailed in Algorithm 1. It  
 296 alternates between the updates of  $u$  and  $v$ , using relaxed fixed point  
 297 iterations associated to (12) and (13). We use the following  $\tau$ -relaxed  
 298 assignment notation

$$299 a \xleftarrow{\tau} b \quad \text{means that} \quad a \leftarrow (1 - \tau)a + \tau b. \quad (14)$$

300 The algorithm outputs the scaled kernel  $\gamma_{i,j} = \exp(K_{i,j})$ .

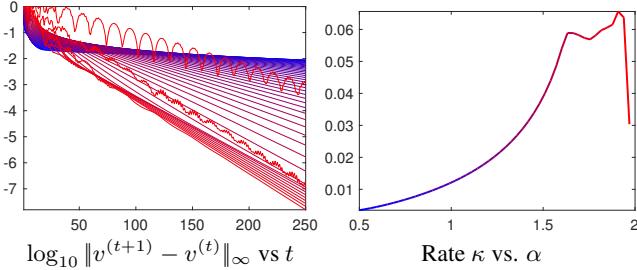
301 **Remark 5** (Choice of  $\tau_k$ ). In the scalar case, i.e.  $d = 1$  (and also  
 302 for isotropic input tensors), when using  $\tau_k = \frac{\varepsilon}{\rho_k + \varepsilon}$  for  $k = 1, 2$ ,  
 303 one retrieves exactly Sinkhorn iterations for unbalanced transport  
 304 as described in [Chizat et al. 2016a], and each update of  $u$  (resp.  
 305  $v$ ) exactly solves the fixed point (12) (resp. (13)). Moreover, it is  
 306 simple to check that these iterates are contractant whenever

$$307 \tau_k \in ]0, \frac{2\varepsilon}{\varepsilon + \rho_k}[ \quad \text{for } k = 1, 2.$$

308 and this property has been observed experimentally for higher dimen-  
 309 sions  $d = 2, 3$ . Using higher values for  $\tau_k$  actually often improves  
 310 the (linear) convergence rate. Figure 3 displays a typical example of  
 311 convergence, and exemplifies the usefulness of using large values of  
 312  $\tau_k$ .

312 **Remark 6** (Stability). In contrast to the usual implementation of  
 313 Sinkhorn’s algorithm, which is numerically unstable for small  $\varepsilon$   
 314 because it requires to compute  $e^{u/\varepsilon}$  and  $e^{v/\varepsilon}$ , the proposed iterations  
 315 using the LSE operator are stable. The algorithm can thus be run  
 316 for arbitrary small  $\varepsilon$ , although the linear speed of convergence is of  
 317 course impacted.

318 **Remark 7** (log and exp computations). A major computational work-  
 319 load of the Q-Sinkhorn Algorithm (1) is the repetitive computation  
 320 of matrix exp and log. For  $d \in \{2, 3\}$  it is possible to use closed  
 321 form expressions to diagonalize the tensors, so that the overall com-  
 322 plexity is comparable with the usual scalar case  $d = 1$ . While the  
 323 applications Section 5 only considers these low-dimensional settings,  
 324 high dimensional problems are of interest, typically for machine  
 325 learning applications. In these cases, one has to resort to iterative  
 326 procedures, such as rapidly converging squaring schemes [Al-Mohy  
 327 and Higham 2009; Al-Mohy and Higham 2012].



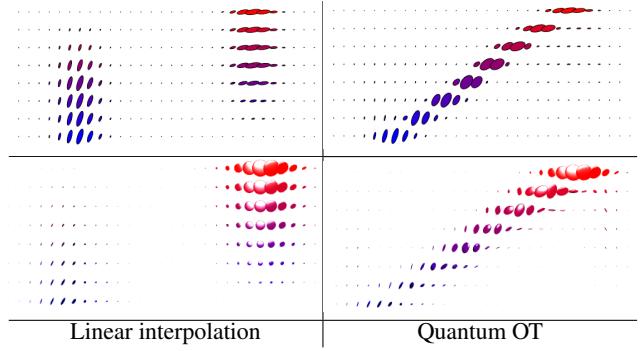
**Figure 3:** Display of convergence of Sinkhorn Algorithm 1 for the example displayed on the first row of Figure 1. Denoting  $u^{(t)}$  the value of the variable  $u$  at iteration  $t$ , the left plot shows the fixed point residual error for increasing values of  $\tau_1 = \tau_2 = \frac{\alpha\varepsilon}{\varepsilon+\rho}$  with  $\alpha \in [0.5, 2]$  (blue to red). The algorithm exhibits a linear convergence rate,  $\log_{10} \|v^{(t+1)} - v^{(t)}\|_\infty \sim -\kappa t$  for some  $\kappa > 0$ , and the right plot displays  $\kappa$  as a function of  $\alpha$ .

```

function QUANTUM-SINKHORN( $\mu, \nu, c, \varepsilon, \rho_1, \rho_2$ )
   $\forall k = 1, 2, \dots, \tau_k \in ]0, \frac{2\varepsilon}{\varepsilon+\rho_k}[$ ,
   $\forall (i, j) \in I \times J, (u_i, v_j) \leftarrow (0_{d \times d}, 0_{d \times d})$ 
  for  $s = 1, 2, 3, \dots$ 
     $K \leftarrow \mathcal{K}(u, v)$ 
     $\forall i \in I, u_i \xleftarrow{\tau_1} \text{LSE}_j(K_{i,j}) - \log(\mu_i)$ 
     $K \leftarrow \mathcal{K}(u, v)$ 
     $\forall j \in J, v_j \xleftarrow{\tau_2} \text{LSE}_i(K_{i,j}) - \log(\nu_j)$ 
  return  $(\gamma_{i,j} = \exp(K_{i,j}))_{i,j}$ 

```

**Algorithm 1:** Quantum-Sinkhorn iterations to compute the optimal coupling  $\gamma$  of the regularized transportation problem (7). The operator  $\mathcal{K}$  is defined in (9).



**Figure 4:** Comparison of linear and quantum-OT interpolation (using formula (6)). Each row shows a field of tensors (top  $d = 2$ , bottom  $d = 3$ ) along a linear segment from  $t = 0$  to  $t = 1$  ( $t$  axis is vertical).

352 tensors.

353 Figure 1 shows larger scale examples. The first row corresponds to  
 354  $X = Y = [0, 1]^2$  and  $d = 2$ , with cost  $c_{i,j} = \|x_i - y_j\|^2 \text{Id}_{2 \times 2}$ ,  
 355 which is a typical setup for image processing. The second row  
 356 corresponds to  $X = Y$  being a triangulated mesh of a surface, and  
 357 the cost is proportional to the squared geodesic distance  $c_{i,j} =$   
 358  $d_X(x_i, y_j)^2 \text{Id}_{2 \times 2}$ .

## 359 4 Quantum Barycenters

360 Following Aguech and Carlier [2011] (see also [Benamou et al. 2015;  
 361 Solomon et al. 2015] for numerical methods using entropic regularization), we now propose a generalization of the OT problem (4),  
 362 where, instead of coupling only two input measures, one tries to  
 363 couple an arbitrary set of inputs, and compute their Fréchet means.  
 364

### 365 4.1 Barycenter Optimization Problem

366 Given some input measures  $(\mu^\ell)_\ell$ , the quantum barycenter problem  
 367 reads

$$\min_{\nu} \sum_{\ell} w_{\ell} W_{\varepsilon}(\mu^{\ell}, \nu), \quad (15)$$

368 where  $(w_{\ell})_{\ell}$  is a set of positive weights normalized so that  $\sum_{\ell} w_{\ell} =$   
 369 1. In the following, for simplicity, we set

$$370 \rho_1 = \rho \quad \text{and} \quad \rho_2 = +\infty$$

371 in the definition (4) of  $W_{\varepsilon}$ . Note that the choice  $\rho_2 = +\infty$  corre-

372 sponds to imposing the exact hard marginal constraint  $\gamma^{\top} \mathbb{1}_J = \nu$ .  
 373 **Remark 9** (Barycenters between single Dirac masses). If all the  
 374 input measures are concentrated on single Diracs  $\mu^{\ell} = P_{\ell} \delta_{x_{\ell}}$ , then  
 375 the single Dirac barycenter (unregularized, i.e.,  $\varepsilon = 0$ ) for a cost  
 376  $d_X(x, y)^{\alpha} \text{Id}_{d \times d}$  is  $P \delta_x^*$  where  $x^* \in X$  is the usual barycenter for  
 the distance  $d_X$ , solving

$$377 x^* \in \operatorname{argmin}_x \mathcal{E}(x) = \sum_{\ell} w_{\ell} d_X^{\alpha}(x_{\ell}, x)$$

378 and the barycentric matrix is

$$379 P = e^{-\frac{\mathcal{E}(x^*)}{\rho}} \exp \left( \sum_{\ell} w_{\ell} \log(P_{\ell}) \right). \quad (16)$$

380 Figure 5 illustrates the effect of a pointwise interpolation (i.e. at the  
 381 same location  $x_{\ell}$  for all  $\ell$ ) between tensors.



**Figure 5:** Two examples of pointwise (without transportation) interpolations (16). Here  $P_1$  and  $P_2$  are represented using the blue/red ellipses on the left/right, and weights are  $(w_1, w_2) = (1 - t, t)$  for  $t \in [0, 1]$  from left to right.

```

function QUANTUM-BARYCENTER( $(\mu_\ell)_{\ell=1}^L, c, \varepsilon, \rho$ )
    Choose  $\tau_1 \in ]0, \frac{2\varepsilon}{\varepsilon+\rho}[$ ,  $\tau_2 \in ]0, 2[$ .
     $\forall (i, j) \in I \times J, (u_i, v_j) \leftarrow (0_{d \times d}, 0_{d \times d})$ 
    for  $s = 1, 2, 3, \dots$ 
        for  $\ell = 1, \dots, L$ 
             $K^\ell \leftarrow \mathcal{K}(u^\ell, v^\ell),$ 
             $\forall i \in I, u_i^\ell \xleftarrow{\tau_1} \text{LSE}_j(K_{i,j}^\ell) - \log(\mu_i^\ell),$ 
             $K^\ell \leftarrow \mathcal{K}(u^\ell, v^\ell).$ 
             $\forall j \in J, \log(\nu_j) \leftarrow \sum_\ell w_\ell (\text{LSE}_i(K_{i,j}^\ell) + v_j^\ell / \varepsilon).$ 
        for  $\ell = 1, \dots, L$ 
             $\forall j \in J, v_j^\ell \xleftarrow{\tau_2} \varepsilon \text{LSE}_i(K_{i,j}^\ell) + v_j^\ell - \varepsilon \log(\nu_j).$ 
    return  $\nu$ 

```

**Algorithm 2:** Quantum-Barycenter iterations to compute the optimal barycenter measure  $\nu$  solving (15). The operator  $\mathcal{K}$  is defined in (18).

380 Problem (15) is convex, and similarly to (8), it can be rewritten in  
 381 dual form.

382 **Proposition 3.** The optimal  $\nu$  solving (15) is solution of

$$\max_{(u^\ell, v^\ell)} \min_{\nu} - \sum_\ell w_\ell \text{tr} \left[ \rho \sum_i e^{u_i^\ell + \log(\mu_i^\ell)} + \sum_j \nu_j v_j^\ell + \varepsilon \sum_{i,j} e^{\mathcal{K}(u^\ell, v^\ell)_{i,j}} \right], \quad (17)$$

383 where here we define  $\mathcal{K}$  as

$$\mathcal{K}(u, v)_{i,j} \stackrel{\text{def.}}{=} -\frac{c_{i,j} + \rho u_i + v_j}{\varepsilon}. \quad (18)$$

## 4.2 Quantum Barycenter Sinkhorn

385 Similarly to Proposition 2, the dual solutions of (17) satisfy a set of  
 386 coupled fixed point equations:

387 **Proposition 4.** Optimal  $(u^\ell, v^\ell)_\ell$  for (17) satisfy

$$\forall (i, \ell), \text{LSE}_j(\mathcal{K}(u^\ell, v^\ell)_{i,j}) - \log(\mu_i^\ell) = u_i^\ell \quad (19)$$

$$\forall (j, \ell), \text{LSE}_i(\mathcal{K}(u^\ell, v^\ell)_{i,j}) = \log(\nu_j) \quad (20)$$

$$\sum_\ell w_\ell v^\ell = 0. \quad (21)$$

388 *Proof.* The proof of (19) and (20) is the same as the one of Proposition-  
 389 2. Minimization of (17) on  $\nu$  leads to (21).  $\square$

390 The extension of the quantum Sinkhorn algorithm to solve the  
 391 barycenter problem (2) is detailed in Algorithm 2. It alternates  
 392 between the updates of  $u$ ,  $\nu$  and  $v$ , using the relaxed version of the  
 393 fixed point equations (19), (20) and (21). The notation  $\xleftarrow{\tau}$  refers to a  
 394 relaxed assignment as defined in (14).

395 **Remark 10 (Choice of  $\tau$ ).** Remark 5 also applies for this Sinkhorn-  
 396 like scheme, and setting  $(\tau_1, \tau_2) = (\frac{\varepsilon}{\rho+\varepsilon}, 1)$  leads, in the scalar case  
 397  $d = 1$ , to the algorithm in [Chizat et al. 2016a]. We found exper-  
 398 imentally that this choice leads to contracting (and hence linearly  
 399 converging) iterations, and that higher values of  $\tau$  usually accelerate  
 400 the convergence rate.

401 **Remark 11 (Scalar and isotropic cases).** Note that in the scalar case  
 402  $d = 1$  and for isotropic input tensors (multiples of the identity), one  
 403 retrieves the provably convergent unbalanced barycenter algorithm  
 404 in [Chizat et al. 2016a].

## 4.3 Numerical Illustrations

406 Figure 6 shows examples of barycenters  $\nu$  solving (15) between four  
 407 input measures  $(\mu_\ell)_{\ell=1}^4$ . The horizontal/vertical axes of the figures  
 408 are indexed by  $(t_1, t_2) \in [0, 1]^2$  (on a  $5 \times 5$  grids) and parameterize  
 409 the weights  $(w_\ell)_{\ell=1}^4$  appearing in (15) as

$$(w_1, w_2, w_3, w_4) \stackrel{\text{def.}}{=} ((1-t_1)(1-t_2), (1-t_1)t_2, t_1(1-t_2), t_1, t_2). \quad (22)$$

410 The left part of Figure 6 corresponds to measures on  $X = Y =$   
 411  $[0, 1]^2$  with  $d = 2$  and ground cost  $c_{i,j} = \|x_i - x_j\|^2 \text{Id}_{2 \times 2}$ . The  
 412 right part of Figure 6 corresponds to measures on  $X = Y$  being  
 413 a surface mesh with  $d = 2$  (the tensors are defined on the tangent  
 414 planes) and a ground cost is  $c_{i,j} = d_X(x_i, x_j)^2 \text{Id}_{2 \times 2}$  where  $d_X$  is  
 415 the geodesic distance on the mesh.

## 5 Applications

417 This section showcases four different applications of Q-OT to regis-  
 418 ter and interpolate tensor fields. Unless otherwise stated, the data is  
 419 normalized to the unit cube  $[0, 1]^d$  (here  $d = 2$  for images) and dis-  
 420 cretized on grids of  $N = 80^d$  points. The regularization parameter  
 421 is set to  $\varepsilon = 0.08^2$ , the fidelity penalty to  $\rho = 1$ , and the relaxation  
 422 parameter for Sinkhorn to  $\tau_k = \frac{1.8\varepsilon}{\varepsilon + \rho_k}$ .

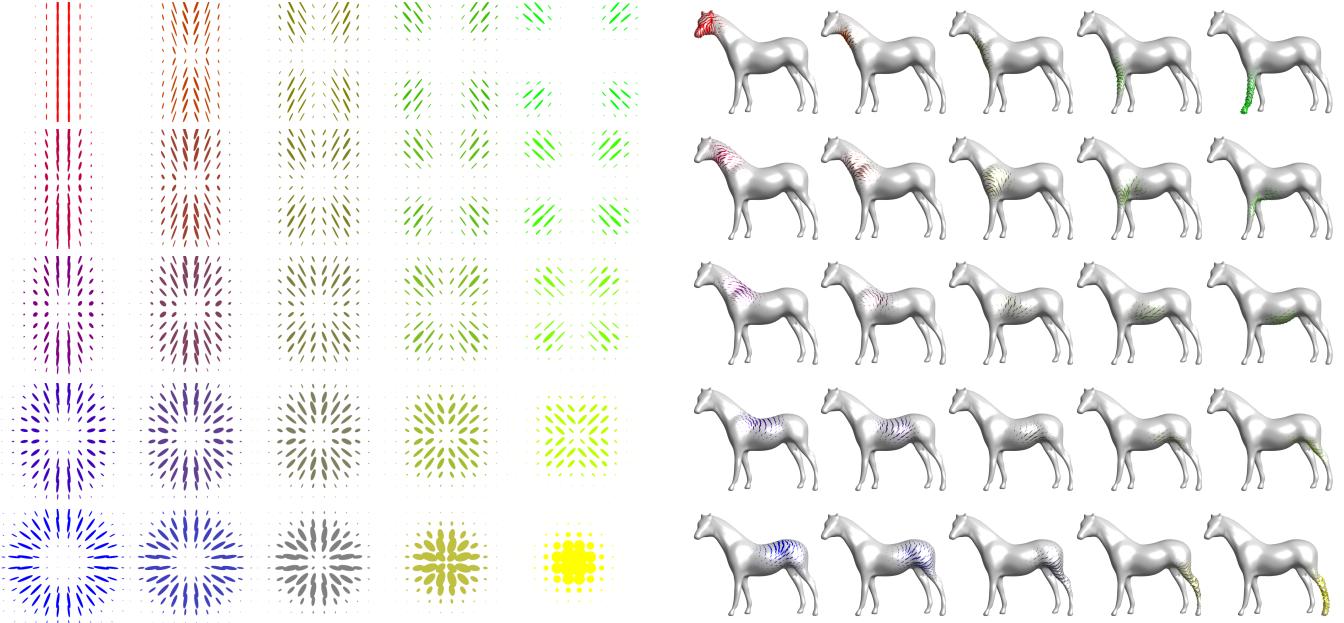
### 5.1 Anisotropic Space-Varying Procedural Noise

424 Texture synthesis using procedural noise functions is widely used  
 425 in rendering pipelines and video games because of both its low  
 426 storage cost and the fact that it is typically parameterized by a few  
 427 meaningful parameters [Lagae et al. 2010]. Following Lagae et  
 428 al. [2011] we consider here a spatially-varying Gabor noise function  
 429 (i.e. non-stationary Gaussian noise), whose covariance function is pa-  
 430 rameterized using a PSD-valued field  $\mu$ . Quantum optimal transport  
 431 allows to interpolate and navigate between these noise functions by  
 432 transporting the corresponding tensor fields. The initial Gabor noise  
 433 method makes use of sparse Gabor splattering [Lagae et al. 2010]  
 434 (which enables synthesis at arbitrary resolution and zooming). For  
 435 simplicity, we rather consider here a more straightforward method,  
 436 where the texture  $f_{t_0}$  is obtained by stopping at time  $t = t_0$  an  
 437 anisotropic diffusion guided by the tensor field  $\mu$  of a high frequency  
 438 noise  $\mathcal{N}$  (numerically a white noise on a grid)

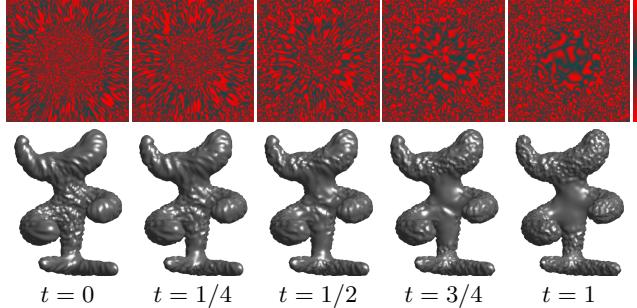
$$\frac{\partial_t f_t}{\partial t} = \text{div}(\mu \nabla f_t), \quad \text{where } f_{t=0} \sim \mathcal{N},$$

439 where  $(\mu \nabla f_t)(x) \stackrel{\text{def.}}{=} \mu(x)(\nabla f_t(x))$  is the vector field obtained by  
 440 applying the tensor  $\mu(x) \in \mathcal{S}_2^+$  to the gradient vector  $\nabla f_t(x) \in \mathbb{R}^2$ .  
 441 Locally around  $x$ , the texture is stretched in the direction of the  
 442 main eigenvector of  $\mu(x)$ , highly anisotropic tensor giving rise to  
 443 elongated “stripes” as opposed to isotropic tensor generating “spots.”

444 Numerically,  $f$  is discretized on a 2-D grid, and  $\mu$  is represented on  
 445 this grid as a sum of Dirac masses (3). On Euclidian domains  $X$ ,  $\nabla$   
 446 and  $\text{div}$  are computed using finite differences, while on triangulated



**Figure 6:**  $5 \times 5$  barycenters of four input measures (displayed in the four corners). The weights  $w \in \mathbb{R}^4$  corresponds to bilinear interpolation weights (22) inside the square.



**Figure 7:** Example of interpolation between two input procedural anisotropic noise function. The PSD tensor field parameterizing the texture are displayed on Figure 1. The colormap used to render the anisotropic texture is displayed on the last column.

mesh, they are implemented using standard piecewise-linear finite element primitives. Figure 7 shows two illustrations of this method. The top row generates an animated color texture by indexing a non-linear black-red colormap (displayed on the right) using  $f_t$ . Bottom row generates an animated bump-mapped surface using  $f_t$  to offset the mesh surface in the normal direction.

## 5.2 Anisotropic Meshing

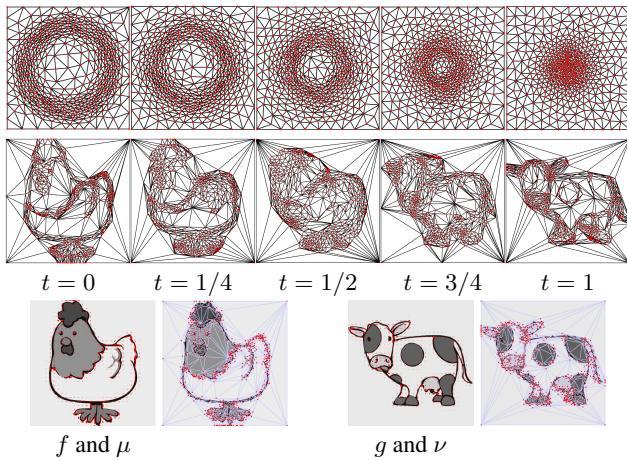
Approximation with anisotropic piecewise linear finite elements on a triangulated mesh is a fundamental tool to address tasks such as discretizing partial differential equations, performing surface remeshing [Alliez et al. 2003] and image compression [Demaret et al. 2006]. A common practice is to generate triangulations complying with a PSD tensor field  $\mu$ , i.e. such that a triangle centered at  $x$  should be inscribed in the ellipsoid  $\{u ; (u - x)^\top \mu(x)(u - x) \leq \delta\}$  for some  $\delta$  controlling the triangulation density. A well-known result is that, to locally approximate a smooth convex  $C^2$  function  $f$ , the optimal shapes of triangles is dictated by the Hessian  $Hf$  of the function

(see [Shewchuk 2002]). In practice, people use  $\mu(x) = |Hf(x)|^\alpha$  for some exponent  $\alpha > 0$  (which is related to the quality measure of the approximation), where  $|\cdot|^\alpha$  indicates the spectral application of the exponentiation (as for matrix exp or log). Figure (8) shows that Q-OT can be used (using formula (6)) to interpolate between two sizing fields  $(\mu, \nu)$ , which are computed from the Hessians (with here  $\alpha = 1$ ) of two initial input images  $(f, g)$ . The resulting anisotropic triangulations are defined as geodesic Delaunay triangulations for the Riemannian metric defined by the tensor field, and are computed using the method detailed in [Bougleux et al. 2009]. This interpolation could typically be used to track the evolution of the solution of some PDE.

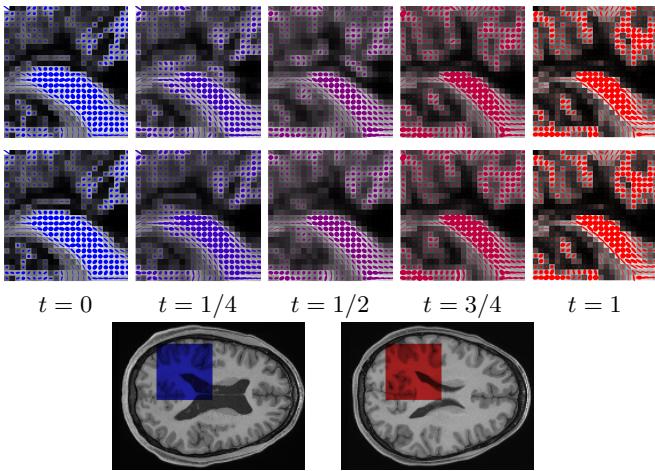
## 5.3 Diffusion Tensor Imaging

Diffusion tensor magnetic resonance imaging (DTI) is a popular technique to image the white matter of the brain (see [Wandell 2016] for a recent overview). DTI measures the diffusion of water molecules, which can be compactly encoded using a PSD tensor field  $\mu(x) \in \mathcal{S}_+^3$ , whose anisotropy and size matches the local diffusivity. A typical goal of this imaging technique is to map the brain anatomical connectivity, and in particular track the white matter fibers. This requires a careful handling of the tensor's energy (its trace) and anisotropy, so that using Q-OT is a perfect fit for such data.

Figure 10 shows an application of Q-OT for the interpolation (using 6) between 2-D slices from DTI tensor fields  $(\mu, \nu)$  acquired on two different subjects. This data is extracted from the studies [Pestilli et al. 2014; Takemura et al. 2016]. These two patients exhibit different anatomical connectivity geometries, and Q-OT is able to track the variation in both orientation and magnitude of the diffusion tensors. This figure also compares the different data fidelity parameters  $\tau \in \{0.05, 1\}$ . Selecting  $\tau = 1$  enforces an overly-strong conservation constraint and leads to interpolation artifacts (in particular some structure are split during the interpolation). In contrast, selecting  $\tau = 0.05$  introduces enough mass creation/destruction during the interpolation to be able to cope with strong inter-subject variability.



**Figure 8:** Two examples of interpolation between two input sizing fields  $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$ . **First row:** triangulation evolution for the sizing fields displayed on Figure 1. **Second row:** the input sizing fields  $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$  are displayed on the third row, and are defined using the absolute valued ( $\alpha = 1$ ) of the Hessian of the underlying images ( $f, g$ ).



**Figure 9:** Interpolation between two 2-D slices of 3-D DTI tensor fields  $(\mu, \nu) = (\mu_{t=0}, \mu_{t=1})$ . For readability, only the X/Y components of the tensors are displayed. **First row:** interpolation obtained using  $\rho = 1$ . **Second row:** interpolation obtained using  $\rho = 0.5$ . **Third row:** anatomical MRI images indicating the region of interest where the computations are performed.

#### 549 5.4 Spectral Color Texture Synthesis

500 As advocated initially in [Galerne et al. 2011], a specific class of  
 501 textured images (so-called micro-textures) is well-modeled using  
 502 stationary Gaussian fields. In the following, we denote  $p$  the pixel positions  
 503 and  $x$  the Fourier frequency indices. For color images, these fields are fully  
 504 characterized by their mean  $m \in \mathbb{R}^3$  and their Fourier power spectrum,  
 505 which is a tensor valued field  $\mu(x)$  where, for each frequency  $x$  (defined on a 2-D grid)  
 506  $\mu(x) \in \mathbb{C}^{3 \times 3}$  is a complex positive semi-definite hermitian matrix. In practice,  
 507  $\mu(x)$  is estimated from an exemplar color image  $f(p) \in \mathbb{R}^3$  using an  
 508 empirical spectrogram  
 509

$$\mu(x) \stackrel{\text{def.}}{=} \frac{1}{K} \sum_{k=1}^K \hat{f}_k(x) \hat{f}_k(x)^* \in \mathbb{C}^{3 \times 3} \quad (23)$$

510 where  $\hat{f}_k$  is the Fourier transform of  $f_k(p) \stackrel{\text{def.}}{=} f(p)w_k(p)$  (computed  
 511 using the FFT),  $w_k$  are windowing functions centred around  $K$  loca-  
 512 tions in the image plane, and  $v^* \in \mathbb{C}^{1 \times 3}$  denoted the transpose-  
 513 conjugate of a vector  $v \in \mathbb{C}^{3 \times 1}$ . Increasing the number  $K$  of  
 514 windowed estimations helps to avoid having rank-deficient covari-  
 515 ances ( $K = 1$  leads to a field  $\mu$  of rank-1 tensors). Randomized  
 516 new textures are then generated by generating random samples  
 517  $F(p) \in \mathbb{R}^3$  from the Gaussian field, which is achieved by defining  
 518 the Fourier transform  $\hat{F}(x) \stackrel{\text{def.}}{=} \hat{N}(x)\sqrt{\mu(x)}\mathbb{1}_3$ , where  $N(p)$  is the  
 519 realization of a Gaussian white noise, and  $\sqrt{\cdot}$  is the matrix square  
 520 root (see [Galerne et al. 2011] for more details). Figure 10 shows  
 521 an application where two input power spectra  $(\mu, \nu)$  (computed us-  
 522 ing (23) from two input textures exemplars  $(f, g)$ ) are interpolated  
 523 using (6), and for each interpolation parameter  $t \in [0, 1]$  a new tex-  
 524 ture  $F$  is synthesized and displayed. Note that while the Q-Sinkhorn  
 525 Algorithm 1 is provided for real PSD matrices, it extends verbatim  
 526 to complex positive hermitian matrices (the matrix logarithm and  
 527 exponential being defined the same way as for real matrices).

## 528 6 Conclusion

In this work, we have proposed a new static formulation for OT  
 529 between tensor-valued measures. This formulation is an extension  
 530 of the recently proposed unbalanced formulation of OT. A chief  
 531 advantage of this formulation is that, once coupled with quantum  
 532 entropic regularization, it leads to an effective numerical scheme,  
 533 which is easily extended to the computation of barycenters.  
 534

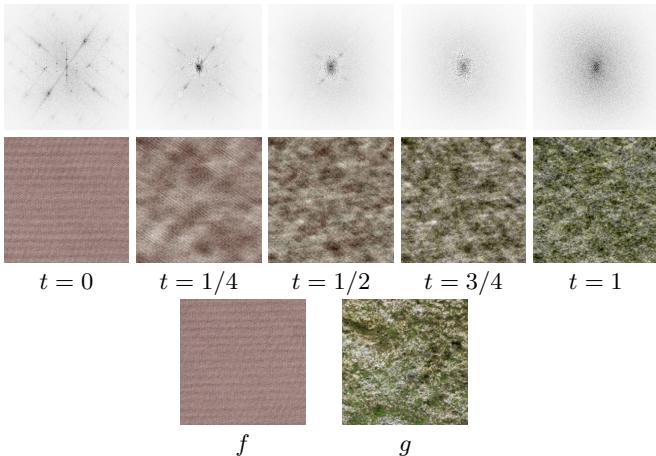
## 535 Acknowledgements

DTI data were provided by Franco Pestilli (NSF IIS 1636893;  
 536 NIH ULTR001108) and the Human Connectome Project (NIH  
 537 1U54MH091657).

Initial results in this paper were shared in a non-peer-reviewed pre-  
 538 publication [Anonymous 2016].

## 541 References

- 542 AGUEH, M., AND CARLIER, G. 2011. Barycenters in the Wasser-  
 543 stein space. *SIAM Journal on Mathematical Analysis* 43, 2, 904–  
 544 924.
- 545 AL-MOHY, A. H., AND HIGHAM, N. J. 2009. A new scaling  
 546 and squaring algorithm for the matrix exponential. *SIAM J. Sci.  
 547 Comput.* 31, 3, 970–989.
- 548 AL-MOHY, A. H., AND HIGHAM, N. J. 2012. Improved inverse  
 549 scaling and squaring algorithms for the matrix logarithm. *SIAM  
 550 J. Sci. Comput.* 34, 4, C153–C169.



**Figure 10:** *First row:* display  $\text{tr}(\mu_t(x))$  where  $\mu_t$  are the interpolated power spectra. *Second row:* realizations of the Gaussian field parameterized by the power spectra  $\mu_t$ . *Third row:* input texture exemplars from which  $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$  are computed.

- 587 CHEN, Y., GEORGIOU, T. T., , AND TANNENBAUM, A. 2016. Distances and riemannian metrics for multivariate spectral densities.  
588 *Preprint arXiv:1610.03041*.
- 589  
590 CHIZAT, L., PEYRÉ, G., SCHMITZER, B., AND VIALARD, F.-X. 2015. Unbalanced optimal transport: Geometry and Kantorovich  
591 formulation. *Preprint 1508.05216, Arxiv*.
- 592  
593 CHIZAT, L., PEYRÉ, G., SCHMITZER, B., AND VIALARD, F.-X. 2016. Scaling algorithms for unbalanced transport problems.  
594 *Preprint 1607.05816, Arxiv*.
- 595  
596 CHIZAT, L., SCHMITZER, B., PEYRÉ, G., AND VIALARD, F.-  
597 X. 2016. An interpolating distance between optimal transport  
598 and Fisher–Rao. *to appear in Foundations of Computational  
599 Mathematics*.
- 600 CRANE, K., DESBRUN, M., AND SCHRÖDER, P. 2010. Trivial  
601 connections on discrete surfaces. In *Computer Graphics Forum*,  
602 vol. 29, Wiley Online Library, 1525–1533.
- 603 CUTURI, M. 2013. Sinkhorn distances: Lightspeed computation of  
604 optimal transportation. In *Proc. NIPS*, vol. 26. 2292–2300.
- 605 DEMARET, L., DYN, N., AND ISKE, A. 2006. Image compression  
606 by linear splines over adaptive triangulations. *Signal Processing*  
607 86, 7, 1604–1616.
- 608 DERICHE, R., TSCHUMPELÉ, D., LENGLET, C., AND ROUSSON,  
609 M. 2006. *Variational Approaches to the Estimation, Regularizatinn and Segmentation of Diffusion Tensor Images*. Springer  
610 US, Boston, MA, 517–530.
- 611 DHILLON, I. S., AND TROPP, J. A. 2008. Matrix nearness problems  
612 with Bregman divergences. *SIAM Journal on Matrix Analysis  
613 and Applications* 29, 4, 1120–1146.
- 614 DRYDEN, I. L., KOLOYDENKO, A., AND ZHOU, D. 2009. Non-  
615 Euclidean statistics for covariance matrices, with applications  
616 to diffusion tensor imaging. *The Annals of Applied Statistics*,  
617 1102–1123.
- 618 DYKSTRA, R. L. 1983. An algorithm for restricted least squares  
619 regression. *J. Amer. Stat.* 78, 384, 839–842.
- 620 FROGNER, C., ZHANG, C., MOBAHI, H., ARAYA, M., AND POG-  
621 GIO, T. 2015. Learning with a Wasserstein loss. In *Advances in  
622 Neural Information Processing Systems*, vol. 28. 2044–2052.
- 623 GALERNE, B., GOUSSEAU, Y., AND MOREL, J.-M. 2011. Random  
624 phase textures: Theory and synthesis. *IEEE Transactions on  
625 image processing* 20, 1, 257–267.
- 626 HOTZ, I., FENG, L., HAGEN, H., HAMANN, B., JOY, K. I., AND  
627 JEREMIC, B. 2004. Physically based methods for tensor field  
628 visualization. IEEE Computer Society, 123–130.
- 629 JIANG, X., NING, L., AND GEORGIOU, T. T. 2012. Distances  
630 and riemannian metrics for multivariate spectral densities. *IEEE  
631 Transactions on Automatic Control* 57, 7 (July), 1723–1735.
- 632 KANTOROVICH, L. 1942. On the transfer of masses (in Russian).  
633 *Doklady Akademii Nauk* 37, 2, 227–229.
- 634 KONDRATYEV, S., MONSAINGEON, L., AND VOROTNIKOV, D.  
635 2015. A new optimal transport distance on the space of finite  
636 Radon measures. Tech. rep., Pre-print.
- 637 KULIS, B., SUSTIK, M. A., AND DHILLON, I. S. 2009. Low-rank  
638 kernel learning with Bregman matrix divergences. *J. Mach. Learn.  
639 Res.* 10 (June), 341–376.

- 641 LAGAE, A., LEFEBVRE, S., COOK, R., DEROSE, T., DRETTAKIS,  
 642 G., EBERT, D. S., LEWIS, J., PERLIN, K., AND ZWICKER,  
 643 M. 2010. A survey of procedural noise functions. In *Computer  
 644 Graphics Forum*, vol. 29, Wiley Online Library, 2579–2600.
- 645 LAGAE, A., LEFEBVRE, S., AND DUTRÉ, P. 2011. Improving  
 646 Gabor noise. *IEEE Transactions on Visualization and Computer  
 647 Graphics* 17, 8, 1096–1107.
- 648 LIERO, M., MIELKE, A., AND SAVARÉ, G. 2015. Optimal entropy-  
 649 transport problems and a new Hellinger–Kantorovich distance  
 650 between positive measures. *ArXiv e-prints*.
- 651 LIERO, M., MIELKE, A., AND SAVARÉ, G. 2016. Optimal transport  
 652 in competition with reaction: The Hellinger–Kantorovich distance  
 653 and geodesic curves. *SIAM Journal on Mathematical Analysis*  
 654 48, 4, 2869–2911.
- 655 MONGE, G. 1781. Mémoire sur la théorie des déblais et des  
 656 remblais. *Histoire de l'Académie Royale des Sciences*, 666–704.
- 657 NING, L., AND GEORGIOU, T. T. 2014. Metrics for matrix-valued  
 658 measures via test functions. In *53rd IEEE Conference on Decision  
 659 and Control*, IEEE, 2642–2647.
- 660 NING, L., GEORGIOU, T. T., AND TANNENBAUM, A. 2015. On  
 661 matrix-valued Monge–Kantorovich optimal mass transport. *IEEE  
 662 transactions on automatic control* 60, 2, 373–382.
- 663 PESTILLI, F., YEATMAN, J. D., ROKEM, A., KAY, K. N., AND  
 664 WANDELL, B. A. 2014. Evaluation and statistical inference for  
 665 human connectomes. *Nature methods* 11, 10, 1058–1063.
- 666 ROCKAFELLAR, R. T. 1970. *Convex analysis*. Princeton University  
 667 Press.
- 668 RUBNER, Y., TOMASI, C., AND GUIBAS, L. J. 2000. The earth  
 669 mover's distance as a metric for image retrieval. *International  
 670 Journal of Computer Vision* 40, 2 (Nov.), 99–121.
- 671 SANTAMBROGIO, F. 2015. Optimal transport for applied mathe-  
 672 maticians. *Progress in Nonlinear Differential Equations and their  
 673 applications* 87.
- 674 SHEWCHEUK, J. 2002. What is a good linear finite element? Interpo-  
 675 lation, conditioning, anisotropy, and quality measures (preprint).  
 676 *University of California at Berkeley* 73.
- 677 SINKHORN, R. 1964. A relationship between arbitrary positive  
 678 matrices and doubly stochastic matrices. *Ann. Math. Statist.* 35,  
 679 876–879.
- 680 SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER,  
 681 A., NGUYEN, A., DU, T., AND GUIBAS, L. 2015. Convolu-  
 682 tional Wasserstein distances: Efficient optimal transportation on  
 683 geometric domains. *TOG* 34, 4 (July), 66:1–66:11.
- 684 TAKEMURA, H., CAIAFA, C. F., WANDELL, B. A., AND PESTILLI,  
 685 F. 2016. Ensemble tractography. *PLoS Comput Biol* 12, 2,  
 686 e1004692.
- 687 VAXMAN, A., CAMPEN, M., DIAMANTI, O., PANIZZO, D.,  
 688 BOMMES, D., HILDEBRANDT, K., AND BEN-CHEN, M. 2016.  
 689 Directional field synthesis, design, and processing. *Comput.  
 690 Graph. Forum* 35, 2, 545–572.
- 691 WANDELL, B. A. 2016. Clarifying human white matter. *Annual  
 692 Review of Neuroscience*.
- 693 WEICKERT, J. 1998. *Anisotropic diffusion in image processing*,  
 694 vol. 1. Teubner Stuttgart.