

Quantum Optimal Transport for Tensor Field Processing

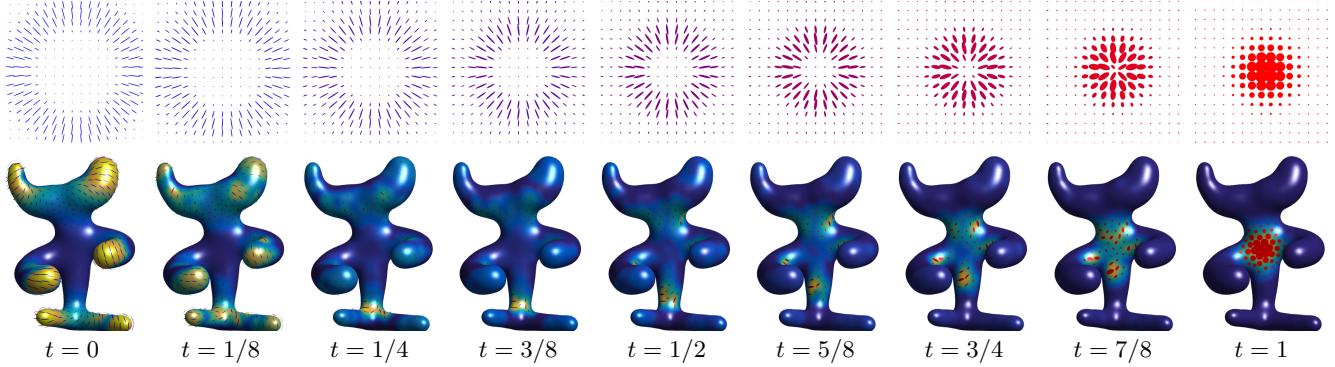


Figure 1: Given two input fields of positive semidefinite matrices (displayed at times $t \in \{0, 1\}$ using ellipses) on some domain (here, a 2-D planar square and a surface mesh), our Quantum Optimal Transport (Q-OT) method defines a continuous interpolating path for $t \in [0, 1]$. Unlike linear 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

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

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

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

Entropic regularization. The current state-of-the-art OT approximation for arbitrary ground costs uses entropic regularization of the transport plan. This leads to strictly convex programs that can be solved using a simple class of highly parallelizable “diagonal scaling” algorithms. The landmark paper of Cuturi [2013] inspired detailed study of these solvers, leading to various generalizations 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) [Wandell 2016] directly maps observed data to fields of tensors, and specific processing methods have been developed (see e.g. [Dryden et al. 2009; Deriche et al. 2006]). Tensor fields are also at the heart of anisotropic diffusions techniques in image processing [Weickert 1998], anisotropic meshing [Alliez et al. 2003; Demaret et al. 2006; Bougleux et al. 2009], anisotropic texture generation [Lagae et al. 2011], 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.¹

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

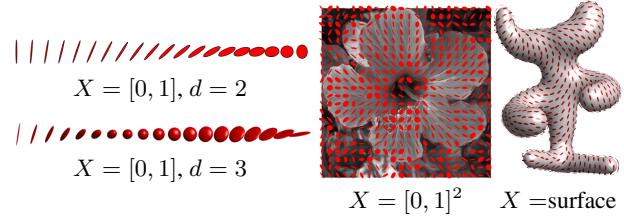


Figure 2: Displays of various types of tensor-valued measures μ . The principal directions of an ellipse at some $x_i \in X$ are the eigenvectors of $\mu_i \in \mathcal{S}_+^d$, while the principal width are given by its eigenvalues.

logarithm $\log(P) = U \text{diag}_s(\log \sigma_s)U^\top$, which is the inverse of \exp on \mathcal{S}_{++}^d .

A tensor-valued measure μ 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 μ 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 μ and the collection 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 collection of matrices.

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

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

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

where ι_C is the indicator function of a closed convex set C , i.e. $\iota_C(P) = 0$ if $P \in C$ and $\iota_C(P) = +\infty$ otherwise. 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 $\xi_i \in \mathcal{S}_{++}^d$, 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

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

¹Available as supplementary material.

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 μ and ν .

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 μ and ν . In the matrix case, $\gamma_{i,j} \in \mathcal{S}_+^d$ is now a PSD matrix, describing how much of mass is moved between μ_i and ν_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 (μ, ν) . But as remarked by Ning et al. [2015], in contrast to the scalar case, in the matrix case (dimension $d > 1$), this constraint is in general 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 (μ, ν) 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 μ_i and ν_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 (μ, ν) 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 (μ, ν) , we denote γ as a solution of (4) or, in practice, its regularized version (see (7) below). The coupling γ 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 illustrations 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 γ to define a path $t \in [0, 1] \mapsto \mu_t$ interpolating between (μ, ν) . 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 $t = 1$).

246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323

Remark 4 (Computational cost). We observed numerically that, similarly to the scalar case, the optimal coupling γ is sparse, meaning that only of the order of $O(|I|)$ non-zero terms are involved in the interpolating measure (6). Note that the entropic regularization algorithm detailed in Section 3 destroys this exact sparsity, but we found numerically that that thresholding to zero the small entries of γ generates accurate approximations.

3 Quantum Sinkhorn

254 The convex program (4) defining quantum OT is computationally 255 challenging because it can be very large scale (problem size is 256 $|I| \times |J|$) for imaging applications, and it involves matrix exponential 257 and logarithm. In this section, leveraging recent advances in 258 computational OT initiated by Cuturi [2013], we propose to use a 259 similar entropy regularized strategy (see also section 1), but this time 260 with the quantum entropy (1).

3.1 Entropic Regularization

261 We define an entropic regularized version of (4)

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

263 Note that when $\varepsilon = 0$, one recovers the original problem (4). This 264 is a strongly convex program, with a unique solution. The crux of 265 this approach, as already known in the scalar case (see [Chizat et al. 266 2016a]), is that its convex dual has a particularly simple structure, 267 which is amenable to a simple alternating maximization strategy.

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

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

269 where $u = (u_i)_{i \in I}, v = (v_j)_{j \in J}$ are collection of symmetric (not 270 necessarily positive) matrices $u_i, v_j \in \mathcal{S}^d$, where we define

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

271 Furthermore, the following primal-dual relationships hold at optimality:

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

273 **Proof.** Applying the Fenchel–Rockafellar duality theorem [Rockafellar 1970] to (7) leads to the dual program

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

275 where here $\text{KL}^*(\cdot | \mu)$ corresponds to the Legendre transform with 276 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 277 Legendre formula leads to the desired result:

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

3.2 Quantum Sinkhorn Algorithm

281 It is possible to use Dykstra's algorithm [1983] (see [Bauschke and 282 Lewis 2000] for its extension to Bregman divergences) to solve (8). 283 This corresponds to alternatively maximizing (8) with respect to u 284 and v . The following proposition states that the maximization with 285 respect to either u or v leads to two fixed-point equations. These 286 fixed points are conveniently written using the log-sum-exp operator, 287

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

288 where the sum on j is replaced by a sum on i for LSE_i .

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

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

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

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

292 **Proof.** Writing the first order condition of (8) with respect to each 293 u_i leads to

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

294 which gives the desired expression. A similar expression holds for 295 the first order conditions with respect to v_j . \square

296 A simple fixed point algorithm is then obtained by replacing in 297 Dykstra's the explicit alternating minimization with respect to u and 298 v by just one step of fixed point iterations (12) and (13). To make 299 the resulting fixed point contractant and ensure linear convergence, 300 one introduces relaxation parameters (τ_1, τ_2) .

301 The quantum Sinkhorn algorithm is detailed in Algorithm 1. It 302 alternates between the updates of u and v , using relaxed fixed point 303 iterations associated to (12) and (13). We use the following τ -relaxed 304 assignment notation

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

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

306 **Remark 5** (Choice of τ_k). In the scalar case, i.e. $d = 1$ (and also 307 for isotropic input tensors), when using $\tau_k = \frac{\varepsilon}{\rho_k + \varepsilon}$ for $k = 1, 2$, 308 one retrieves exactly Sinkhorn iterations for unbalanced transport 309 as described in [Chizat et al. 2016a], and each update of u (resp. 310 v) exactly solves the fixed point (12) (resp. (13)). Moreover, it is 311 simple to check that these iterates are contractant whenever

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

312 and this property has been observed experimentally for higher dimensions 313 $d = 2, 3$. Using higher values for τ_k actually often improves 314 the (linear) convergence rate. Figure 3 displays a typical example of 315 convergence, and exemplifies the usefulness of using large values of 316 τ_k , which leads to a speed-up of a factor 6 with respect to the usual 317 Sinkhorn's choice $\tau_k = \frac{\varepsilon}{\varepsilon + \rho_k}$.

318 **Remark 6** (Stability). In contrast to the usual implementation of 319 Sinkhorn's algorithm, which is numerically unstable for small ε 320 because it requires to compute $e^{u/\varepsilon}$ and $e^{v/\varepsilon}$, the proposed iterations 321 using the LSE operator are stable. The algorithm can thus be run 322 for arbitrary small ε , although the linear speed of convergence is of 323 course impacted.

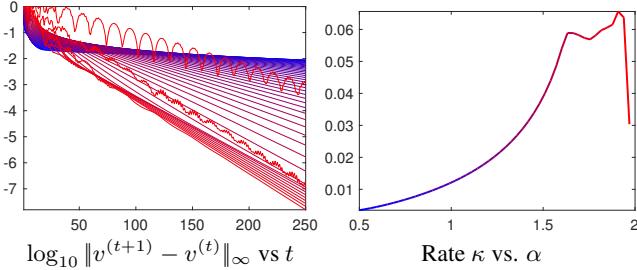


Figure 3: Display of convergence of Sinkhorn Algorithm 1 for the example displayed on the first row of Figure 1. Denoting $u^{(t)}$ the value of the variable u at iteration t , the left plot shows the fixed point residual error for increasing values of $\tau_1 = \tau_2 = \frac{\alpha\varepsilon}{\varepsilon+\rho}$ 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 κ as a function of α .

```

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

```

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

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

Remark 8 (Computational complexity). For low-dimensional problems (typically for those considered in Section 5), the Q-Sinkhorn Algorithm 1 scales to grid sizes of roughly 6k points (with machine-precision solutions computed in a few minutes on a standard laptop). For large scale grids, even storing the full coupling γ becomes prohibitive. We however observed numerically that, similarly to the usual scalar case, the optimal γ solving (7) is highly sparse (up to machine precision for small enough ε). We thus found that using the multi-scale refinement strategy introduced in [Schmitz 2016] is able to make the Q-Sinkhorn scales to high resolution grids. It is not used to produce the figures of this article, but it is available in the companion computational toolbox.

3.3 Numerical Illustrations

Figures 1, 4 and 5 illustrates on synthetic examples of input tensor fields (μ, ν) the Q-OT interpolation method. We recall that it is obtained in two steps:

- One first computes the optimal γ solving (7) using Sinkhorn

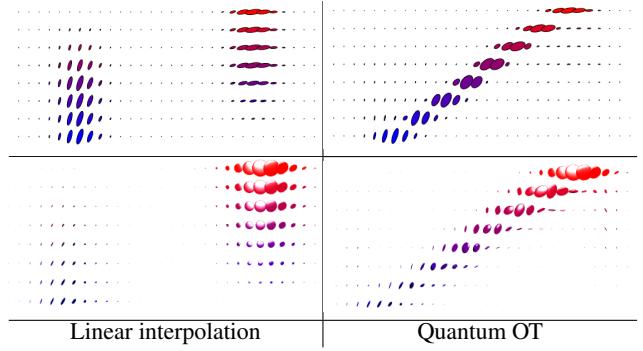


Figure 4: Comparison of linear and quantum-OT interpolation (using formula (6)). Each row shows a tensor field μ_t (top $d = 2$, bottom $d = 3$) along a linear segment from $t = 0$ to $t = 1$ (t axis is vertical).

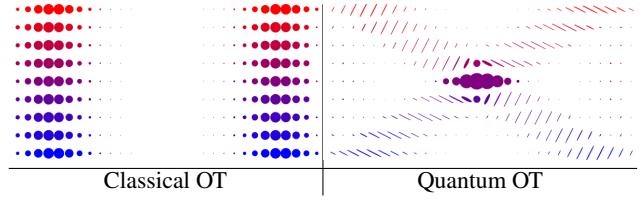


Figure 5: Comparison of classical OT (i.e. between isotropic tensors) and quantum-OT (between anisotropic tensors) interpolation (using the same display as Figure 4).

iterations (Algorithm 1).

- Then, for any $t \in [0, 1]$, one computes μ_t using this optimal γ with formula (6).

Figure 4 shows examples of interpolations on a 1-D domain $X = Y = [0, 1]$ with tensors of dimension $d = 2$ and $d = 3$, and a ground cost $c_{i,j} = |x_i - y_j|^2 \text{Id}_{d \times d}$. It compares the OT interpolation, which achieves a “mass displacement,” to the usual linear interpolation $(1-t)\mu + t\nu$, which only performs a pointwise interpolation of the tensors.

Figure 5 shows the effect of taking into account the anisotropy of tensors into the definition of OT. In the case of isotropic tensors (see Remark 1), the method reduces to the usual scalar OT, and in 1-D it corresponds to the monotone re-arrangement [Santambrogio 2015]. In contrast, the Q-OT of anisotropic tensors is forced to reverse the ordering of the transport map in order for tensors with similar orientations to be matched together.

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

4 Quantum Barycenters

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

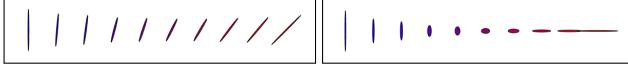


Figure 6: Two examples of pointwise (without transportation) interpolations, using formula (16). Here P_1 and P_2 are represented using the blue/red ellipses on the left/right, and weights are $(w_1, w_2) = (1 - t, t)$ for $t \in [0, 1]$ from left to right.

379 4.1 Barycenter Optimization Problem

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

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

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

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

384 in the definition (4) of W_{ε} . Note that the choice $\rho_2 = +\infty$ corre-
385 sponds to imposing the exact hard marginal constraint $\gamma^{\top} \mathbb{1}_J = \nu$.
386 **Remark 9** (Barycenters between single Dirac masses). If all the
387 input measures are concentrated on single Diracs $\mu^{\ell} = P_{\ell} \delta_{x_{\ell}}$, then
388 the single Dirac barycenter (unregularized, i.e., $\varepsilon = 0$) for a cost
389 $d_X(x, y)^{\alpha} \text{Id}_{d \times d}$ is $P \delta_x^*$ where $x^* \in X$ is the usual barycenter for
390 the distance d_X , solving

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

391 and the barycentric matrix is

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

392 Figure 6 illustrates the effect of a pointwise interpolation (i.e. at the
393 same location x_{ℓ} for all ℓ) between tensors.

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

396 **Proposition 3.** *The optimal ν solving (15) is solution of*

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

397 where here we define \mathcal{K} as

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

398 4.2 Quantum Barycenter Sinkhorn

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

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

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

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

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

402 *Proof.* The proof of (19) and (20) is the same as the one of Proposi-
403 tion 2. Minimization of (17) on ν leads to (21). \square

function QUANTUM-BARYCENTER($(\mu_{\ell})_{\ell=1}^L, c, \varepsilon, \rho$)

Choose $\tau_1 \in]0, \frac{2\varepsilon}{\varepsilon+\rho}[$, $\tau_2 \in]0, 2[$.

$\forall (i, j) \in I \times J, \quad (u_i, v_j) \leftarrow (0_{d \times d}, 0_{d \times d})$

for $s = 1, 2, 3, \dots$

for $\ell = 1, \dots, L$

$K^{\ell} \leftarrow \mathcal{K}(u^{\ell}, v^{\ell})$,

$\forall i \in I, \quad u_i^{\ell} \xleftarrow{\tau_1} \text{LSE}_j(K_{i,j}^{\ell}) - \log(\mu_i^{\ell})$,

$K^{\ell} \leftarrow \mathcal{K}(u^{\ell}, v^{\ell})$.

$\forall j \in J, \quad \log(\nu_j) \leftarrow \sum_{\ell} w_{\ell} (\text{LSE}_i(K_{i,j}^{\ell}) + v_j^{\ell}/\varepsilon)$.

for $\ell = 1, \dots, L$

$\forall j \in J, \quad v_j^{\ell} \xleftarrow{\tau_2} \varepsilon \text{LSE}_i(K_{i,j}^{\ell}) + v_j^{\ell} - \varepsilon \log(\nu_j)$.

return ν

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

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

409 **Remark 10** (Choice of τ). Remark 5 also applies for this Sinkhorn-
410 like scheme, and setting $(\tau_1, \tau_2) = (\frac{\varepsilon}{\rho+\varepsilon}, 1)$ leads, in the scalar case
411 $d = 1$, to the algorithm in [Chizat et al. 2016a]. We found experimen-
412 tally that this choice leads to contracting (and hence linearly
413 converging) iterations, and that higher values of τ usually accelerate
414 the convergence rate.

415 **Remark 11** (Scalar and isotropic cases). Note that in the scalar case
416 $d = 1$ and for isotropic input tensors (multiples of the identity), one
417 retrieves the provably convergent unbalanced barycenter algorithm
418 in [Chizat et al. 2016a].

419 4.3 Numerical Illustrations

420 Figure 7 shows examples of barycenters ν solving (15) between four
421 input measures $(\mu_{\ell})_{\ell=1}^4$. The horizontal/vertical axes of the figures
422 are indexed by $(t_1, t_2) \in [0, 1]^2$ (on a 5×5 grids) and parameterize
423 the weights $(w_{\ell})_{\ell=1}^4$ appearing in (15) as

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

424 The left part of Figure 7 corresponds to measures on $X = Y =$
425 $[0, 1]^2$ with $d = 2$ and ground cost $c_{i,j} = \|x_i - x_j\|^2 \text{Id}_{2 \times 2}$. The
426 right part of Figure 7 corresponds to measures on $X = Y$ being
427 a surface mesh with $d = 2$ (the tensors are defined on the tangent
428 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
429 the geodesic distance on the mesh.

5 Applications

430 This section showcases four different applications of Q-OT to regis-
431 ter and interpolate tensor fields. Unless otherwise stated, the data
432 is normalized to the unit cube $[0, 1]^d$ (here $d = 2$ for images) and
433 discretized on grids of $|I| = |J| = 60^d$ points. The regularization
434 parameter is set to $\varepsilon = 0.08^2$, the fidelity penalty to $\rho = 1$, and the
435 relaxation parameter for Sinkhorn to $\tau_k = \frac{1.8\varepsilon}{\varepsilon+\rho k}$.

437 5.1 Anisotropic Space-Varying Procedural Noise

438 Texture synthesis using procedural noise functions is widely used
439 in rendering pipelines and video games because of both its low

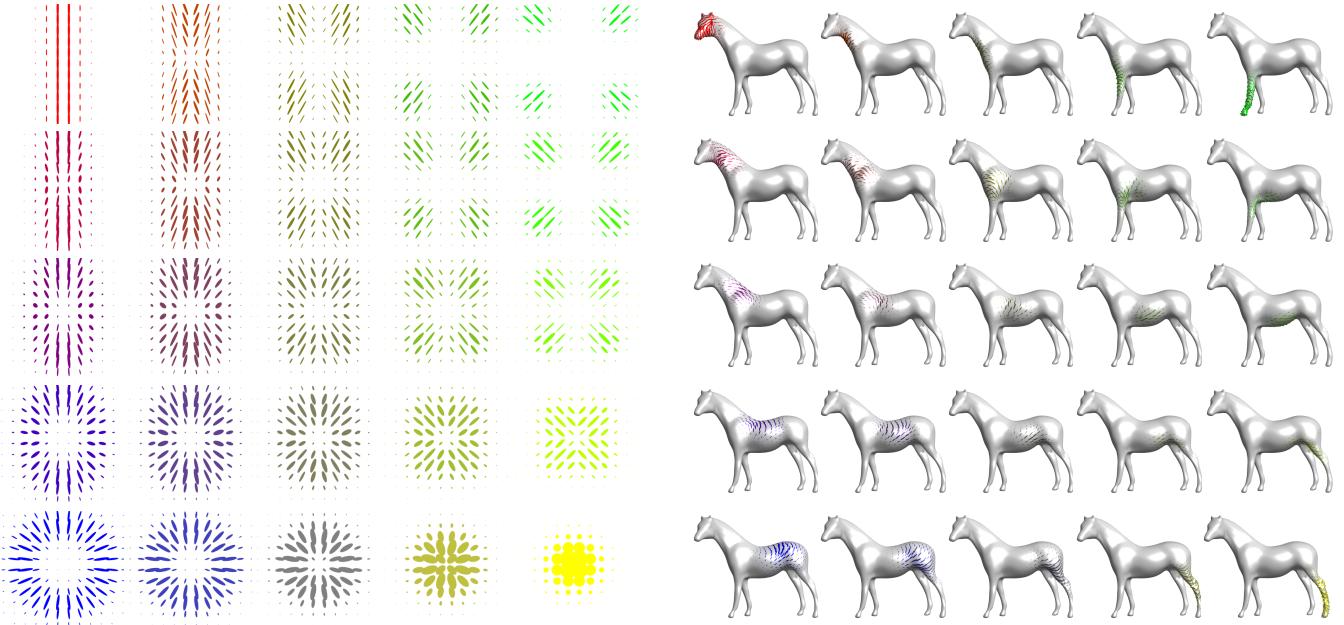


Figure 7: 5×5 barycenters of four input measures (displayed in the four corners). The weights $w \in \mathbb{R}^4$ correspond to bilinear interpolation weights (22) inside the square.

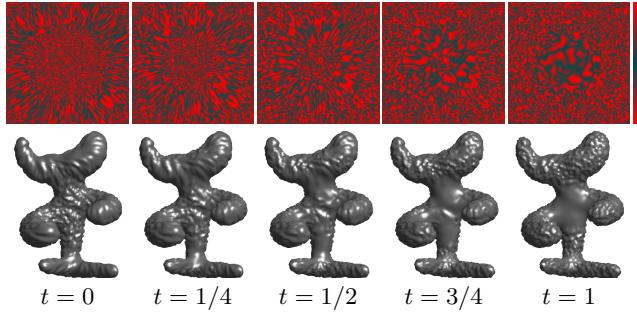


Figure 8: Example of interpolation between two input procedural anisotropic noise function. The PSD tensor field parameterizing the texture are displayed on Figure 1. The colormap used to render the anisotropic texture is displayed on the last column.

Locally around x , the texture is stretched in the direction of the main eigenvector of $\mu(x)$, highly anisotropic tensor giving rise to elongated “stripes” as opposed to isotropic tensor generating “spots.”

Numerically, f is discretized on a 2-D grid, and μ is represented on this grid as a sum of Dirac masses (3). On Euclidian domains X , ∇ and div are computed using finite differences, while on triangulated mesh, they are implemented using standard piecewise-linear finite element primitives. Figure 8 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 sizing field μ , i.e. such that a triangle centered at $x \in X$ should be inscribed in the ellipsoid $\{u \in X ; (u - x)^\top \mu(x)(u - x) \leq \delta\}$ for some δ 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 9 shows that Q-OT can be used (using formula (6)) to interpolate between two sizing fields (μ, ν) , which are computed from the Hessians (with here $\alpha = 1$) of two initial input images (f, g) . The resulting anisotropic triangulations are computed using the method detailed in [Bougleux et al. 2009]. They corresponds to geodesic Delaunay triangulations for the Riemannian metric defined by the

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

where $(\mu \nabla f_t)(x) \stackrel{\text{def}}{=} \mu(x)(\nabla f_t(x))$ is the vector field obtained by applying the tensor $\mu(x) \in \mathcal{S}_2^+$ to the gradient vector $\nabla f_t(x) \in \mathbb{R}^2$.

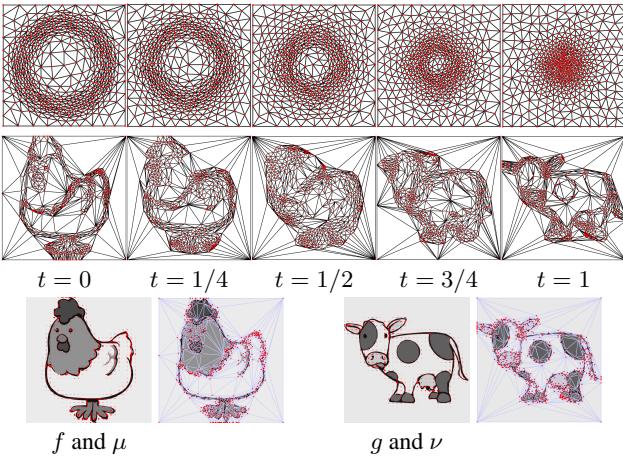


Figure 9: Two examples of interpolation between two input sizing fields $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$. **First row:** triangulation evolution for the sizing fields displayed on Figure 1. **Second row:** the input sizing fields $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$ are displayed on the third row, and are defined using the absolute value ($\alpha = 1$) of the Hessian of the underlying images (f, g).

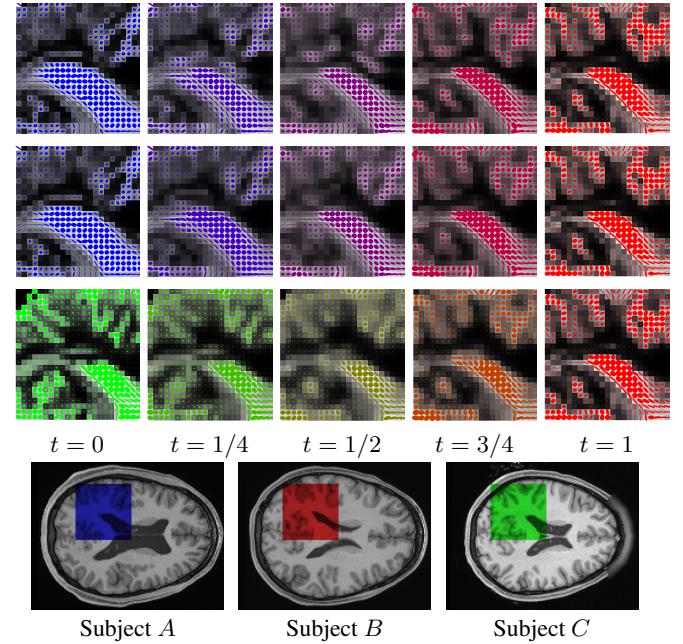


Figure 10: Interpolation between two 2-D slices of 3-D DTI tensor fields $(\mu, \nu) = (\mu_{t=0}, \mu_{t=1})$. For readability, only the X/Y components of the tensors are displayed. **First row:** interpolation between subjects (A, B) obtained using $\rho = 1$. **Second row:** interpolation between subjects (A, B) obtained using $\rho = 0.05$. **Third row:** interpolation between subjects (C, B) obtained using $\rho = 0.05$. **Fourth row:** anatomical MRI images of subjects (A, B, C) indicating the region of interest where the computations are performed.

489 tensor field. This interpolation could typically be used to track the
490 evolution of the solution of some PDE.

491

5.3 Diffusion Tensor Imaging

492 Diffusion tensor magnetic resonance imaging (DTI) is a popular
493 technique to image the white matter of the brain (see [Wandell
494 2016] for a recent overview). DTI measures the diffusion of water
495 molecules, which can be compactly encoded using a PSD tensor
496 field $\mu(x) \in \mathcal{S}_+^3$, whose anisotropy and size matches the local
497 diffusivity. A typical goal of this imaging technique is to map
498 the brain anatomical connectivity, and in particular track the white
499 matter fibers. This requires a careful handling of the tensor's energy
500 (its trace) and anisotropy, so that using Q-OT is a perfect fit for such
501 data.

502 Figure 10 shows an application of Q-OT for the interpolation (us-
503 ing 6) between 2-D slices from DTI tensor fields (μ, ν) acquired on
504 two different subjects. This data is extracted from the studies [Pestilli
505 et al. 2014; Takemura et al. 2016]. These two patients exhibit differ-
506 ent anatomical connectivity geometries, and Q-OT is able to track
507 the variation in both orientation and magnitude of the diffusion ten-
508 sors. This figure also compares the different data fidelity parameters
509 $\rho \in \{0.05, 1\}$. Selecting $\rho = 1$ enforces an overly-strong conserva-
510 tion constraint and leads to interpolation artifacts (in particular some
511 structure are split during the interpolation). In contrast, selecting
512 $\rho = 0.05$ introduces enough mass creation/destruction during the
513 interpolation to be able to cope with strong inter-subject variability.

524

\mathbb{R}^3 using an empirical spectrogram

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

525 where \hat{f}_k is the Fourier transform of $f_k(p) \stackrel{\text{def}}{=} f(p)w_k(p)$ (computed
526 using the FFT), w_k are windowing functions centred around K loc-
527 ations in the image plane, and $v^* \in \mathbb{C}^{1 \times 3}$ denoted the transpose-
528 conjugate of a vector $v \in \mathbb{C}^{3 \times 1}$. Increasing the number K of win-
529 dowed estimations helps to avoid having rank-deficient covariances
530 ($K = 1$ leads to a field μ of rank-1 tensors).

531 Randomized new textures are then created by generating random
532 samples $F(p) \in \mathbb{R}^3$ from the Gaussian field, which is achieved
533 by defining the Fourier transform $\hat{F}(x) \stackrel{\text{def}}{=} m + \hat{N}(x)\sqrt{\mu(x)}\mathbf{1}_3$,
534 where $N(p)$ is the realization of a Gaussian white noise, and $\sqrt{\cdot}$ is
535 the matrix square root (see [Galerne et al. 2011] for more details).

536 Figure 11 shows an application where two input power spectra (μ, ν)
537 (computed using (23) from two input textures exemplars (f, g)) are
538 interpolated using (6), and for each interpolation parameter $t \in [0, 1]$
539 a new texture F is synthesized and displayed. Note that while the
540 Q-Sinkhorn Algorithm 1 is provided for real PSD matrices, it
541 extends verbatim to complex positive hermitian matrices (the matrix
542 logarithm and exponential being defined the same way as for real
543 matrices).

References

545 AGUEH, M., AND CARLIER, G. 2011. Barycenters in the Wasser-
546 stein space. *SIAM Journal on Mathematical Analysis* 43, 2, 904–

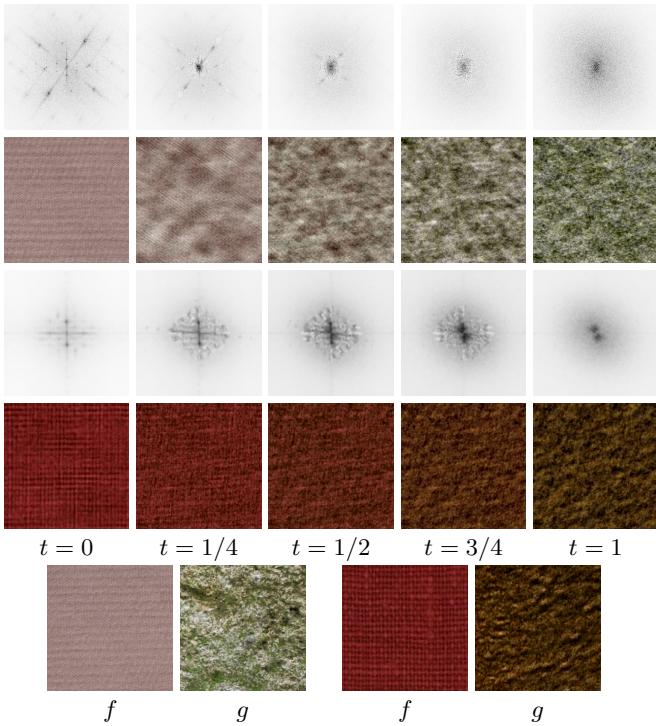


Figure 11: **Row 1 and 3:** display $\text{tr}(\mu_t(x))$ where μ_t are the interpolated power spectra. **Rows 2 and 4:** realizations of the Gaussian field parameterized by the power spectra μ_t . **Row 5:** input texture exemplars from which $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$ are computed.

- 573 BREGMAN, L. M. 1967. The relaxation method of finding the
574 common point of convex sets and its application to the solution of
575 problems in convex programming. *USSR comp. math. and math.*
576 *phys.* 7, 3, 200–217.
- 577 BRENIER, Y. 1991. Polar factorization and monotone rearrangement
578 of vector-valued functions. *Comm. Pure Appl. Math.* 44, 4, 375–
579 417.
- 580 CARLEN, E. A., AND MAAS, J. 2014. An analog of the 2-
581 Wasserstein metric in non-commutative probability under which
582 the fermionic Fokker–Planck equation is gradient flow for the
583 entropy. *Communications in Mathematical Physics* 331, 3, 887–
584 926.
- 585 CHANDRASEKARAN, V., AND SHAH, P. 2016. Relative entropy
586 optimization and its applications. *Mathematical Programming*,
587 1–32.
- 588 CHEN, Y., GEORGIOU, T. T., , AND TANNENBAUM, A. 2016. Dis-
589 tances and Riemannian metrics for multivariate spectral densities.
590 *Preprint arXiv:1610.03041*.
- 591 CHIZAT, L., PEYRÉ, G., SCHMITZER, B., AND VIALARD, F.-X. 2015. Unbalanced optimal transport: Geometry and Kantorovich
592 formulation. *Preprint 1508.05216, Arxiv*.
- 593 CHIZAT, L., PEYRÉ, G., SCHMITZER, B., AND VIALARD, F.-X. 2016. Scaling algorithms for unbalanced transport problems.
594 *Preprint 1607.05816, Arxiv*.
- 595 CHIZAT, L., SCHMITZER, B., PEYRÉ, G., AND VIALARD, F.-X. 2016. An interpolating distance between optimal transport
596 and Fisher–Rao. *to appear in Foundations of Computational
597 Mathematics*.
- 598 CRANE, K., DESBRUN, M., AND SCHRÖDER, P. 2010. Trivial
599 connections on discrete surfaces. In *Computer Graphics Forum*,
600 vol. 29, Wiley Online Library, 1525–1533.
- 601 CUTURI, M. 2013. Sinkhorn distances: Lightspeed computation of
602 optimal transportation. In *Proc. NIPS*, vol. 26. 2292–2300.
- 603 DEMARET, L., DYN, N., AND ISKE, A. 2006. Image compression
604 by linear splines over adaptive triangulations. *Signal Processing*
605 86, 7, 1604–1616.
- 606 DERICHE, R., TSCHUMPELÉ, D., LENGLLET, C., AND ROUSSON,
607 M. 2006. *Variational Approaches to the Estimation, Regular-
608 izatiinn and Segmentation of Diffusion Tensor Images*. Springer
609 US, Boston, MA, 517–530.
- 610 DHILLON, I. S., AND TROPP, J. A. 2008. Matrix nearness problems
611 with Bregman divergences. *SIAM Journal on Matrix Analysis
612 and Applications* 29, 4, 1120–1146.
- 613 DRYDEN, I. L., KOLOYDENKO, A., AND ZHOU, D. 2009. Non-
614 Euclidean statistics for covariance matrices, with applications
615 to diffusion tensor imaging. *The Annals of Applied Statistics*,
616 1102–1123.
- 617 DYKSTRA, R. L. 1983. An algorithm for restricted least squares
618 regression. *J. Amer. Stat.* 78, 384, 839–842.
- 619 FROGNER, C., ZHANG, C., MOBAHI, H., ARAYA, M., AND POG-
620 GIO, T. 2015. Learning with a Wasserstein loss. In *Advances in
621 Neural Information Processing Systems*, vol. 28. 2044–2052.
- 622 GALERNE, B., GOUSSEAU, Y., AND MOREL, J.-M. 2011. Random
623 phase textures: Theory and synthesis. *IEEE Transactions on
624 image processing* 20, 1, 257–267.

- 628 HOTZ, I., FENG, L., HAGEN, H., HAMANN, B., JOY, K. I., AND 683 SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER,
629 JEREMIC, B. 2004. Physically based methods for tensor field 684 A., NGUYEN, A., DU, T., AND GUIBAS, L. 2015. Convolutional
630 visualization. IEEE Computer Society, 123–130. 685 Wasserstein distances: Efficient optimal transportation on
686 geometric domains. *TOG 34*, 4 (July), 66:1–66:11.
- 631 JIANG, X., NING, L., AND GEORGIOU, T. T. 2012. Distances 687 TAKEMURA, H., CAIAFA, C. F., WANDELL, B. A., AND PESTILLI,
632 and Riemannian metrics for multivariate spectral densities. *IEEE 688 F. 2016. Ensemble tractography. *PLoS Comput Biol 12*, 2,*
633 *Transactions on Automatic Control 57*, 7 (July), 1723–1735. 689 e1004692.
- 634 KANTOROVICH, L. 1942. On the transfer of masses (in Russian). 690 VAXMAN, A., CAMPEN, M., DIAMANTI, O., PANZZO, D.,
635 *Doklady Akademii Nauk 37*, 2, 227–229. 691 BOMMES, D., HILDEBRANDT, K., AND BEN-CHEN, M. 2016.
636 KONDRATYEV, S., MONSAINGEON, L., AND VOROTNIKOV, D. 692 Directional field synthesis, design, and processing. *Comput.*
637 2015. A new optimal transport distance on the space of finite 693 *Graph. Forum 35*, 2, 545–572.
- 638 Radon measures. Tech. rep., Pre-print.
- 639 KULIS, B., SUSTIK, M. A., AND DHILLON, I. S. 2009. Low-rank 694 WANDELL, B. A. 2016. Clarifying human white matter. *Annual
640 kernel learning with Bregman matrix divergences. *J. Mach. Learn.* 695 Review of Neuroscience.*
- 641 *Res. 10* (June), 341–376.
- 642 LAGAE, A., LEFEBVRE, S., COOK, R., DEROSSE, T., DRETTAKIS, 696 WEICKERT, J. 1998. *Anisotropic diffusion in image processing*,
643 G., EBERT, D. S., LEWIS, J., PERLIN, K., AND ZWICKER, 697 vol. 1. Teubner Stuttgart.
- 644 M. 2010. A survey of procedural noise functions. In *Computer
645 Graphics Forum*, vol. 29, Wiley Online Library, 2579–2600.
- 646 LAGAE, A., LEFEBVRE, S., AND DUTRÉ, P. 2011. Improving 698 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- 699 MONGE, G. 1781. Mémoire sur la théorie des déblais et des
700 transport problems and a new Hellinger–Kantorovich distance
701 between positive measures. *ArXiv e-prints*.
- 702 LIERO, M., MIELKE, A., AND SAVARÉ, G. 2016. Optimal transport
703 in competition with reaction: The Hellinger–Kantorovich distance
704 and geodesic curves. *SIAM Journal on Mathematical Analysis*
705 48, 4, 2869–2911.
- 706 NING, L., AND GEORGIOU, T. T. 2014. Metrics for matrix-valued 707 NING, L., GEORGIOU, T. T., AND TANNENBAUM, A. 2015. On
708 measures via test functions. In *53rd IEEE Conference on Decision
709 and Control*, IEEE, 2642–2647.
- 710 ROCKAFELLAR, R. T. 1970. *Convex analysis*. Princeton University
711 Press.
- 712 RUBNER, Y., TOMASI, C., AND GUIBAS, L. J. 2000. The earth
713 mover’s distance as a metric for image retrieval. *International
714 Journal of Computer Vision 40*, 2 (Nov.), 99–121.
- 715 SANTAMBROGIO, F. 2015. Optimal transport for applied mathe- 716 SCHMITZER, B. 2016. Stabilized sparse scaling algorithms for
717 maticians. *Progress in Nonlinear Differential Equations and their
718 applications 87*.
- 719 SHEWCHUK, J. 2002. What is a good linear finite element? Interpo-
720 lation, conditioning, anisotropy, and quality measures (preprint).
721 *University of California at Berkeley 73*.
- 722 SINKHORN, R. 1964. A relationship between arbitrary positive
723 matrices and doubly stochastic matrices. *Ann. Math. Statist. 35*,
724 876–879.