

Quantum Optimal Transport for Tensor Field Processing

Gabriel Peyré*
 CNRS
 École Normale Supérieure

Lenaïc Chizat
 Univ. Paris-Dauphine
 INRIA Mokaplan

François-Xavier Vialard
 Univ. Paris-Dauphine
 INRIA Mokaplan

Justin Solomon
 MIT

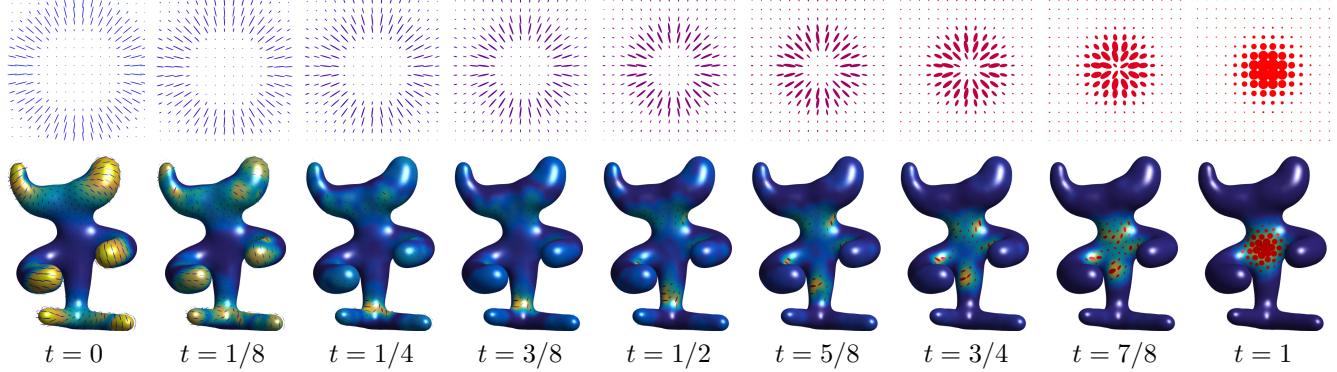


Figure 1: given two input fields of PSD matrices (displayed at time $t \in \{0, 1\}$ using ellipses) on some domain (here a 2-D planar square and a surface mesh), our Quantum Optimal Transport (Q-OT) method defines a continuous interpolating paths for $t \in [0, 1]$. On contrary to usual linear interpolation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

Abstract

This article introduces a new notion of optimal transport (OT) between tensor fields, which are measures whose values are positive semidefinite matrices (PSD). This “quantum” formulation of 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.

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. This OT distance offers many interesting features, and in particular leads to a geometrically faithful way to manipulate and interpolate probability distributions.

*gabriel.peyre@ens.fr

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).

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 [Monge 1781], its reformulation by Kantorovich [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 [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 (Frechet’s mean) of input measures, a convex program studied by Aguech and Carlier [Aguech and Carlier 2011]. OT has found numerous applications, for instance in computer vision (under the naming “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 to approximate OT 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 [Cuturi 2013] inspired detailed study of these solvers, leading to various generalizations of Sinkhorn’s algorithm [Sinkhorn 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 uses 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

¹<https://github.com/gpeyre/2016-wasserstein-tensor-valued>

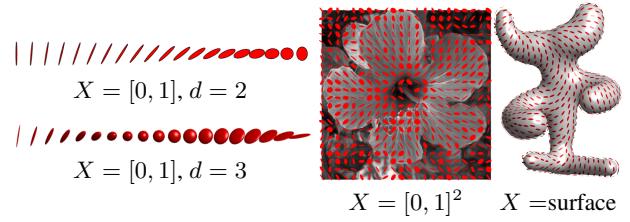


Figure 2: Displays of various types of tensor-valued measures μ . The principal directions of an ellipse at some $x_i \in X$ are the eigenvectors of $\mu_i \in \mathcal{S}_+^d$, while the principal width are given by its eigenvalues.

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 special case of 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. This means that 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 particular 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 vector of matrix $(\mu_i)_i$. 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

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

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 μ and ν .

2.1 Tensor Transportation

Following the initial static formulation of OT by Kantorovich [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 $d > 1$), in general this constraint is too strong, and there might exist no coupling satisfying these marginal constraints. We advocate in this work that the natural workaround for the matrix setting is in fact 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 thus 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 in [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 along each eigendirection.

Remark 2 (Cost between single Dirac masses). When $\mu = P\delta_x$ and $\nu = Q\delta_x$ are two Dirac masses at 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 a solution of (4) (in practice, its regularized version, see (7) below). The coupling γ defines a (fuzzy) correspondance between the tensor fields. A typical use of this correspondance 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 (barycenters) between more than two input fields.

Mimicking the definition of the optimal transport interpolation (the so-called McCann displacement interpolation see for instance [Sanambrogio 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 the 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), and 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's tensor $\gamma_{i,j}^t \delta_{x_i^t}$ 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 $t = 1$).

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

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

3.1 Entropic Regularization

We define an entropic regularized version of (4)

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

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

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

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

where we define

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

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

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

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

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

where here $\text{KL}^*(\cdot | \mu)$ corresponds to the Legendre transform with 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 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

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

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

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

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

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

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

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

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

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

which gives the desired expression. A similar expression holds for the first order conditions with respect to v_j . □

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

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

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

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

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

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

and this property has been observed experimentally for higher dimensions $d = 2, 3$. Using higher values for τ_k actually often improves the (linear) convergence rate; we leave this interesting observation open for future theoretical investigation.

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

Remark 7 (log and exp computations). **[ToDo: explain close forms]**

```

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

```

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

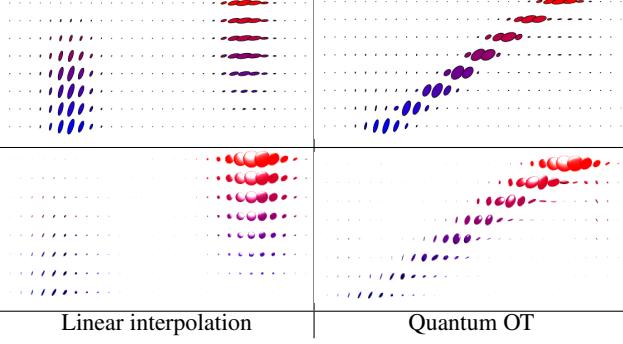


Figure 3: Comparison of linear and quantum-OT interpolation (using formula (6)). Each row shows a field of tensors (top $d = 2$, bottom $d = 3$) along a linear segment from $t = 0$ to $t = 1$ (t axis is vertical).

3.3 Numerical Illustrations

Figures 1 and 3 illustrates on synthetic examples of input tensor fields (μ, ν) our interpolation method. We recall that it is obtained in two steps:

1. One first computes the optimal γ solving (7) using Sinkhorn iterations (Algorithm 1).
2. Then, for any $t \in [0, 1]$, one computes μ_t using this optimal γ with formula (6).

Figure 3 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 ?? shows larger scale examples. The top row corresponds to $X = Y$ being a triangulated mesh of a surface, and the cost is proportional to the squared geodesic distance $c_{i,j} = d_X(x_i, y_j)^2 \text{Id}_{2 \times 2}$. The bottom 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 background shows the output of an anisotropic diffusion driven by the tensor field).

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 Frechet means.

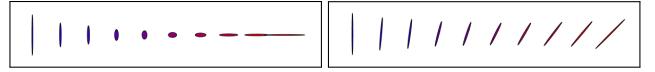


Figure 4: Two examples of pointwise (without transportation) interpolations (16). Here P_1 and P_2 are represented using the blue/red ellipses on the left/right, and weights are $(w_1, w_2) = (1-t, t)$ for $t \in [0, 1]$ from left to right.

4.1 Barycenters Optimization Problem

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

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

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

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

in the definition (4) of W_ε . Note that the choice $\rho_2 = +\infty$ corresponds to imposing the exact hard marginal constraint $\gamma^\top \mathbf{1}_J = \nu$. *Remark 8* (Barycenters between single Dirac masses). If all the input measures are concentrated on single Diracs $\mu^\ell = P_\ell \delta_{x_\ell}$, then the single Dirac barycenter (unregularized, i.e., $\varepsilon = 0$) for a cost $d_X(x, y)^\alpha \text{Id}_{d \times d}$ is $P \delta_x^*$ where $x^* \in X$ is the usual barycenter for the distance d_X , solving

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

and the barycentric matrix is

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

Figure (4) illustrate the effect of a pointwise interpolation (i.e. at the same location x_ℓ for all ℓ) between tensors.

Problem (15) is convex, and similarly to (8), it can be re-written in dual form.

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

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

where here we defined \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

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

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

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

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

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

Proof. The proof of (19) and (20) is the same as the one of Proposition 2. Minimization of (17) on ν leads to (21). \square

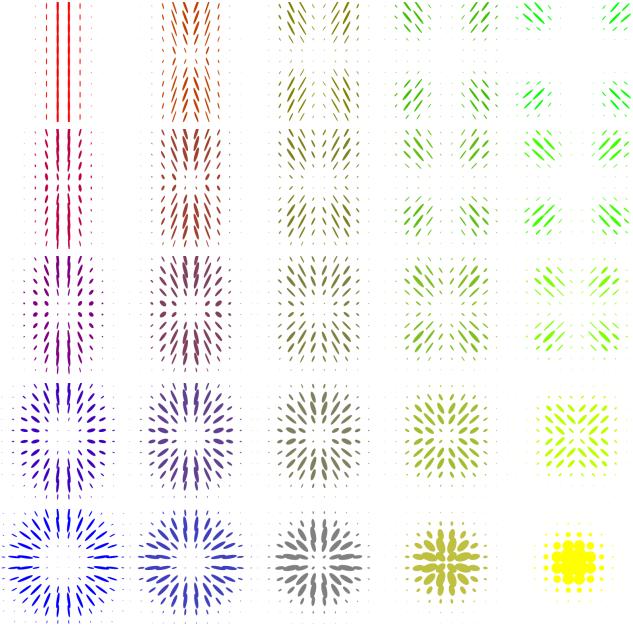


Figure 5: 5×5 barycenters of four input measures (displayed in the four corners). The weights $w \in \mathbb{R}^4$ corresponds to bilinear interpolation weights (22) inside the square.

```

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  $\nu$ 

```

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

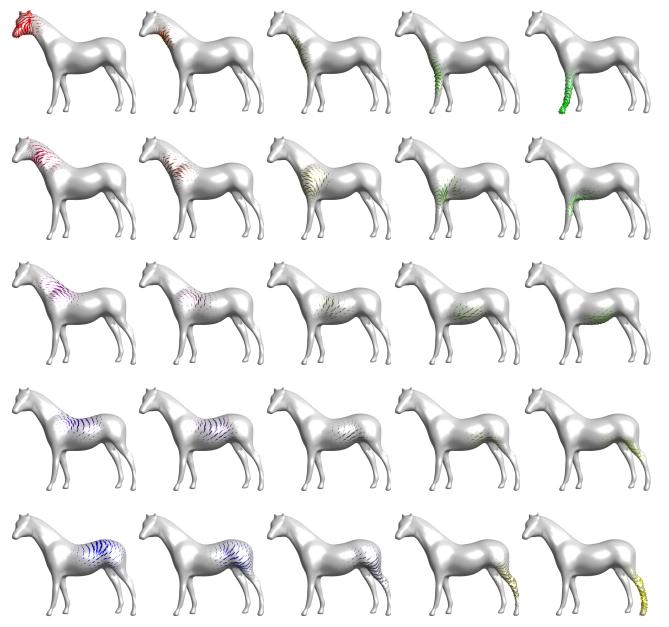
The extension of the quantum Sinkhorn algorithm to solve the barycenter problem (2) is detailed in Algorithm 2. It alternate between the update of u , ν and v , using relaxed version of the fixed point equations (19), (20) and (21). The notation $\xleftarrow{\tau}$ refers to a relaxed assignment as defined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

Remark 9 (Choice of τ). Remarks 5 also applies for this Sinkhorn-like scheme, and setting $(\tau_1, \tau_2) = (\frac{\varepsilon}{\rho+\varepsilon}, 1)$ leads, in the scalar case $d = 1$, to the algorithm detailed in [Chizat et al. 2016a]. We found experimentally that this choice leads to contracting (and hence linearly converging) iterations, and that higher values of τ usually accelerates the convergence rate.

Remark 10 (Scalar and isotropic cases). Note that in the scalar case $d = 1$ and for isotropic input tensors (multiples of the identity), one retrieves the provably convergent unbalanced barycenter algorithm detailed in [Chizat et al. 2016a].

4.3 Numerical Illustrations

Figure ?? and ?? shows examples of barycenters ν solving (15) between four input measures $(\mu_\ell)_{\ell=1}^4$. The horizontal/vertical axes



of the figures are indexed by $(t_1, t_2) \in [0, 1]^2$ (on a 5×5 grids) and parameterize the weights $(w_\ell)_{\ell=1}^4$ appearing in (15) as

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

Figure ?? corresponds to measures on $X = Y = [0, 1]^2$ with $d = 2$ and ground cost $c_{i,j} = \|x_i - x_j\|^2 \text{Id}_{2 \times 2}$. Figure ?? corresponds to measures on $X = Y$ being a surface mesh with $d = 2$ (the tensors are defined on the tangent 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 the geodesic distance on the mesh.

5 Applications

5.1 Anisotropic Space-varying Procedural Noise

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

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

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 $\nabla f_t(x) \in \mathbb{R}^2$. 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”.

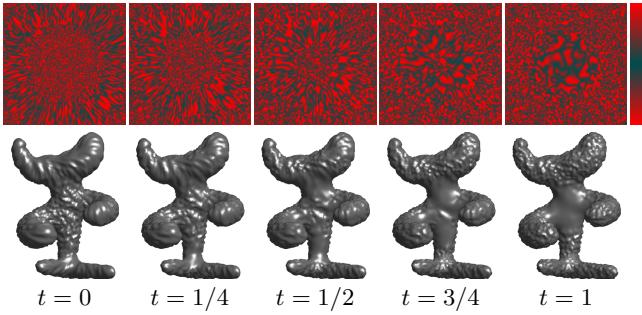


Figure 6: Example of interpolation between two input procedural anisotropic noise function. The PSD tensor field parameterizing the texture are displayed on Figure 1. The colormap used to render the anisotropic texture is displayed on the last column.

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 domain, ∇ and div are computed using usual finite difference, while on triangulated mesh, they are implemented using standard piecewise linear finite element primitives. Figure 6 shows two illustration of this method. Top row generates an animated color texture by indexing a non-linear black-red color map 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] or for image compression [Demaret et al. 2006]. A common practice is to generate triangulation complying with a PSD tensor field μ , i.e. such that a triangle centred at x should be inscribed in the ellipsoid $\{u ; (u - x)^\top \mu(x)(u - x) \leq \delta\}$ for some δ controlling the triangulation density. A well known result is that optimal triangle shapes to locally approximate a smooth convex C^2 function f is dictated by the Hessian Hf of the function [Shewchuk 2002], so that in practice people use $\mu(x) = |Hf(x)|^\alpha$ for some exponent $\alpha > 0$ (which is related to the quality measure of the approximation). Figure (7) shows that Q-OT can be used (using formula (6)) to interpolate between two sizing fields (μ, ν) , which are computed from the Hessian of two initial input images (f_0, f_1) . The resulting anisotropic triangulations are defined as geodesic Delaunay triangulation for the Riemannian metric defined by the tensor field, and computed using the method detailed in [Bougleux et al. 2009]. This interpolation could typically used to track the evolution of the solution of some PDE.

5.3 Diffusion Tensor Imaging

5.4 Spectral Color Texture Synthesis

Note that while the Q-Sinkhorn Algorithm 1 is detailed for real PSD matrices, it extend verbatim to complex positive hermitian matrices (the matrix logarithm and exponential being defined the same way as for real matrices).

6 Conclusion

In this work, we have proposed a new static formulation for OT between tensor-valued measures. This formulation is an extension of the recently proposed unbalanced formulation of OT. A chief advantage of this formulation is that, once coupled with quantum

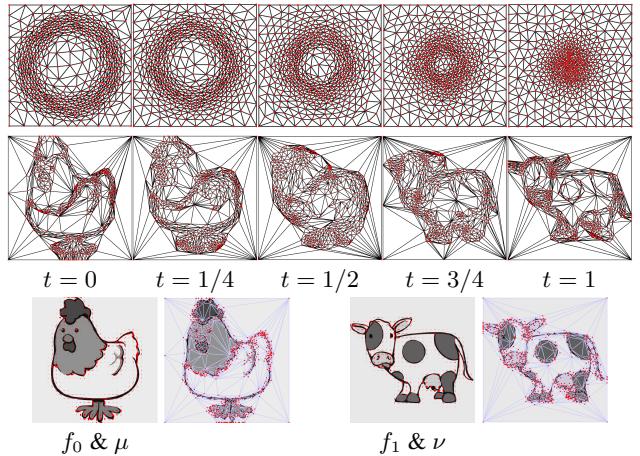


Figure 7: Two examples of interpolation between two input sizing fields (μ, ν) . **First row:** the sizing fields are displayed on Figure 1. **Second row:** the sizing fields (μ, ν) are displayed on the last row, and are defined using the absolute valued of the Hessian of the underlying images (f_0, f_1) .

entropic regularization, it leads to an effective numerical scheme, which is easily extended to the computation of barycenters.

Acknowledgements

The work of Gabriel Peyré has been supported by the European Research Council (ERC project SIGMA-Vision).

References

- AGUEH, M., AND CARLIER, G. 2011. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis* 43, 2, 904–924.
- ALLIEZ, P., COHEN-STEINER, D., DEVILLERS, O., LÉVY, B., AND DESBRUN, M. 2003. Anisotropic polygonal remeshing. In *ACM Transactions on Graphics (TOG)*, vol. 22, ACM, 485–493.
- BAUSCHKE, H. H., AND LEWIS, A. S. 2000. Dykstra’s algorithm with Bregman projections: a convergence proof. *Optimization* 48, 4, 409–427.
- BENAMOU, J.-D., AND BRENIER, Y. 2000. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* 84, 3, 375–393.
- BENAMOU, J.-D., CARLIER, G., CUTURI, M., NENNA, L., AND PEYRÉ, G. 2015. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comp.* 37, 2, A1111–A1138.
- BONNEEL, N., VAN DE PANNE, M., PARIS, S., AND HEIDRICH, W. 2011. Displacement interpolation using Lagrangian mass transport. *ACM Trans. Graph.* 30, 6 (Dec.), 158:1–158:12.
- BOUGLEUX, S., PEYRÉ, G., AND COHEN, L. D. 2009. Image compression with anisotropic geodesic triangulations. In *Proc. of ICCV’09*, 2343–2348.
- BREGMAN, L. M. 1967. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR comp. math. and math. phys.* 7, 3, 200–217.

- BRENIER, Y. 1991. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* 44, 4, 375–417.
- CARLEN, E. A., AND MAAS, J. 2014. An analog of the 2-wasserstein metric in non-commutative probability under which the fermionic fokker–planck equation is gradient flow for the entropy. *Communications in Mathematical Physics* 331, 3, 887–926.
- CHANDRASEKARAN, V., AND SHAH, P. 2016. Relative entropy optimization and its applications. *Mathematical Programming*, 1–32.
- CHEN, Y., GEORGIOU, T. T., , AND TANNENBAUM, A. 2016. Distances and riemannian metrics for multivariate spectral densities. *Preprint arXiv:1610.03041*.
- CHIZAT, L., PEYRÉ, G., SCHMITZER, B., AND VIALARD, F.-X. 2015. Unbalanced optimal transport: Geometry and kantorovich formulation. *Preprint 1508.05216*, Arxiv.
- CHIZAT, L., PEYRÉ, G., SCHMITZER, B., AND VIALARD, F.-X. 2016. Scaling algorithms for unbalanced transport problems. *Preprint 1607.05816*, Arxiv.
- CHIZAT, L., SCHMITZER, B., PEYRÉ, G., AND VIALARD, F.-X. 2016. An interpolating distance between optimal transport and Fisher-Rao. *to appear in Foundations of Computational Mathematics*.
- CRANE, K., DESBRUN, M., AND SCHRÖDER, P. 2010. Trivial connections on discrete surfaces. In *Computer Graphics Forum*, vol. 29, Wiley Online Library, 1525–1533.
- CUTURI, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transportation. In *Proc. NIPS*, vol. 26, 2292–2300.
- DEMARET, L., DYN, N., AND ISKE, A. 2006. Image compression by linear splines over adaptive triangulations. *Signal Processing* 86, 7, 1604–1616.
- DERICHE, R., TSCHUMPELÉ, D., LENGLER, C., AND ROUSSON, M. 2006. *Variational Approaches to the Estimation, Regularization and Segmentation of Diffusion Tensor Images*. Springer US, Boston, MA, 517–530.
- DHILLON, I. S., AND TROPP, J. A. 2008. Matrix nearness problems with bregman divergences. *SIAM Journal on Matrix Analysis and Applications* 29, 4, 1120–1146.
- DRYDEN, I. L., KOLOYDENKO, A., AND ZHOU, D. 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 1102–1123.
- DYKSTRA, R. L. 1983. An algorithm for restricted least squares regression. *J. Amer. Stat.* 78, 384, 839–842.
- FROGNER, C., ZHANG, C., MOBAHI, H., ARAYA, M., AND POGGIO, T. 2015. Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems*, vol. 28, 2044–2052.
- HOTZ, I., FENG, L., HAGEN, H., HAMANN, B., JOY, K. I., AND JEREMIC, B. 2004. Physically based methods for tensor field visualization. IEEE Computer Society, 123–130.
- JIANG, X., NING, L., AND GEORGIOU, T. T. 2012. Distances and riemannian metrics for multivariate spectral densities. *IEEE Transactions on Automatic Control* 57, 7 (July), 1723–1735.
- KANTOROVICH, L. 1942. On the transfer of masses (in russian). *Doklady Akademii Nauk* 37, 2, 227–229.
- KONDRATYEV, S., MONSAINGEON, L., AND VOROTNIKOV, D. 2015. A new optimal transport distance on the space of finite Radon measures. *Tech. rep., Pre-print*.
- KULIS, B., SUSTIK, M. A., AND DHILLON, I. S. 2009. Low-rank kernel learning with bregman matrix divergences. *J. Mach. Learn. Res.* 10 (June), 341–376.
- LAGAE, A., LEFEBVRE, S., COOK, R., DEROSSE, T., DRETTAKIS, G., EBERT, D. S., LEWIS, J., PERLIN, K., AND ZWICKER, M. 2010. A survey of procedural noise functions. In *Computer Graphics Forum*, vol. 29, Wiley Online Library, 2579–2600.
- LAGAE, A., LEFEBVRE, S., AND DUTRÉ, P. 2011. Improving gabor noise. *IEEE Transactions on Visualization and Computer Graphics* 17, 8, 1096–1107.
- LIERO, M., MIELKE, A., AND SAVARÉ, G. 2015. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *ArXiv e-prints*.
- LIERO, M., MIELKE, A., AND SAVARÉ, G. 2016. Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis* 48, 4, 2869–2911.
- MONGE, G. 1781. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*, 666–704.
- NING, L., AND GEORGIOU, T. T. 2014. Metrics for matrix-valued measures via test functions. In *53rd IEEE Conference on Decision and Control*, IEEE, 2642–2647.
- NING, L., GEORGIOU, T. T., AND TANNENBAUM, A. 2015. On matrix-valued monge–kantorovich optimal mass transport. *IEEE transactions on automatic control* 60, 2, 373–382.
- ROCKAFELLAR, R. T. 1970. *Convex analysis*. Princeton university press.
- RUBNER, Y., TOMASI, C., AND GUIBAS, L. J. 2000. The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision* 40, 2 (Nov.), 99–121.
- SANTAMBROGIO, F. 2015. Optimal transport for applied mathematicians. *Progress in Nonlinear Differential Equations and their applications* 87.
- SHEWCHUK, J. 2002. What is a good linear finite element? interpolation, conditioning, anisotropy, and quality measures (preprint). *University of California at Berkeley* 73.
- SINKHORN, R. 1964. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.* 35, 876–879.
- SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T., AND GUIBAS, L. 2015. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *TOG* 34, 4 (July), 66:1–66:11.
- VAXMAN, A., CAMPEN, M., DIAMANTI, O., PANZZO, D., BOMMES, D., HILDEBRANDT, K., AND BEN-CHEN, M. 2016. Directional field synthesis, design, and processing. *Comput. Graph. Forum* 35, 2, 545–572.
- WEICKERT, J. 1998. *Anisotropic diffusion in image processing*, vol. 1. Teubner Stuttgart.