

Quantum entropic regularization of matrix-valued optimal transport†

GABRIEL PEYRÉ¹, LÉNAÏC CHIZAT²,
FRANÇOIS-XAVIER VIALARD² and JUSTIN SOLOMON³

¹*CNRS and DMA, École Normale Supérieure, 45 rue d'Ulm - F 75230 PARIS cedex 05*
email: gabriel.peyre@ens.fr

²*Ceremade, Univ. Paris-Dauphine and INRIA Mokaplan, Place du Maréchal de Lattre de Tassigny, 75016 Paris*
email: chizat@ceremade.dauphine.fr, vialard@ceremade.dauphine.fr

³*EECS and CSAIL, MIT, 32 Vassar Street, room 32-D460 Cambridge, MA 02139*
email: jsolomon@mit.edu

(Received 11 April 2017; revised 6 September 2017; accepted 6 September 2017; first published online 28 September 2017)

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 optimal transport (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 barycentres 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.

Key words: Optimal transport, tensor field, PSD matrices, quantum entropy.

1 Introduction

Optimal transport (OT) is an active field of research at the intersection of probability theory, Partial Differential Equations (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. This paper proposes a computationally tractable way to extend OT to matrix-valued measures. Figure 1 showcase a typical application of this method to the interpolation of tensor fields.

† The work of Gabriel Peyré has been supported by the European Research Council (ERC project SIGMA-Vision). J. Solomon acknowledges support of Army Research Office grant W911NF-12-R-0011 (“Smooth Modeling of Flows on Graphs”).

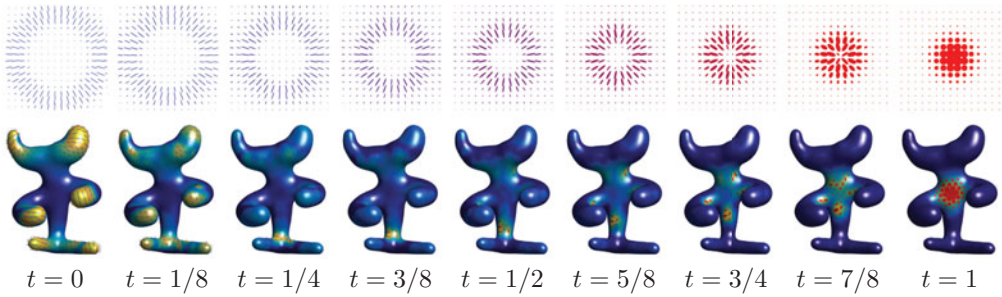


FIGURE 1. Given two input fields of positive semidefinite matrices (displayed at times $t \in \{0, 1\}$ using ellipses) on some domain (here, a 2-D planar square and a surface mesh), our Quantum Optimal Transport (Q-OT) method defines a continuous interpolating path for $t \in [0, 1]$. Unlike linear 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 barycentre of several input fields, is computed using a fast extension of the well-known Sinkhorn algorithm.

1.1 Previous work

1.1.1 Scalar-valued optimal transport

Dating back to the eighteenth century, classical instances of the OT 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 [44], its reformulation by Kantorovich [36] 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 [13] and opened the door to a dynamical (geodesic) reformulation [8]. We refer to [52] 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 barycentre (Fréchet mean) of input measures, a convex program studied by Agueh and Carlier [1]. OT also has found numerous applications, for instance in computer vision (under the name “earth mover’s distance”) [51] and computer graphics [10].

1.1.2 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 [21, 37, 42] and a static one [20, 41] and has been applied in machine learning [30]. Our work extends this static unbalanced formulation to tensor-valued measures.

1.1.3 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

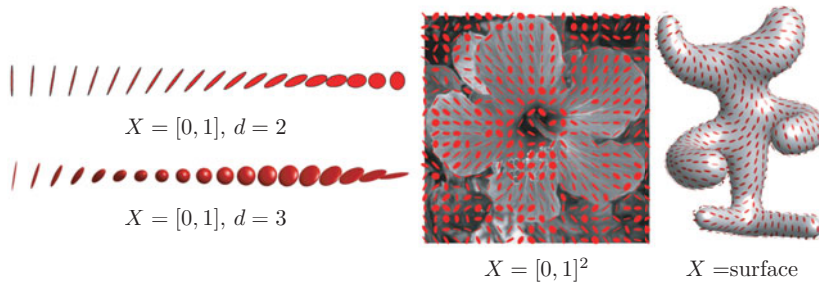


FIGURE 2. Displays of various types of tensor-valued measures μ . The principal directions of an ellipse at some $x_i \in X$ are the eigenvectors of $\mu_i \in \mathcal{S}_+^d$, while the principal widths are given by its eigenvalues.

be solved using a simple class of highly parallelizable diagonal scaling algorithms. The landmark paper of Cuturi [24] inspired detailed study of these solvers, leading to various generalizations of Sinkhorn’s algorithm [55]. This includes for instance the use of fast convolutional structures [56], extensions to barycentres [9] and application to unbalanced OT [22, 30]. 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 [12] and Dykstra’s algorithm [6, 29]. 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 [16, 27, 38].

1.1.4 Tensor field processing

Tensor-valued data are ubiquitous in various areas of imaging science, computer graphics and vision. Figure 2 displays a few examples of matrix-valued tensor fields. In medical imaging, diffusion tensor imaging (DTI) [59] directly maps observed data to fields of tensors, and specific processing methods have been developed for this class of data (see e.g. [26, 28]). Tensor fields are also at the heart of anisotropic diffusion techniques in image processing [60], anisotropic meshing [4, 11, 25] and anisotropic texture generation [40]; they also find application in line drawing [58] and data visualization [34].

1.1.5 OT and Sinkhorn on pairs of tensors

Although this is not the topic of this paper, we note that several notions of OT have been defined between two tensors *without* any spatial displacement. Gurvits introduced in [33] a Sinkhorn-like algorithm to couple two tensors by an endomorphism that preserves positivity. This algorithm, however, is only known to converge in the case where the two involved tensors are the identity matrices; see [32] for a detailed analysis. In contrast to this “static” formulation that seeks for a joint coupling between the tensors, a geodesic dynamic formulation is proposed in [14]; see also [15, 17, 18, 43] for related approaches.

1.1.6 OT on tensor fields

The simplest way to define OT-like distances between arbitrary vector-valued measures is to use dual norms [46], which correspond to generalizations of W_1 OT for which transport cost equals ground distance. The corresponding metrics, however, have degenerate behaviour in interpolation and barycentre 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 [18,35], see also [19] with applications to colour image processing. A static formulation of a tensor-valued OT is proposed in [47], 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 barycentres 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 (PSD) 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 $\exp(P) = U \operatorname{diag}_s(e^{\sigma_s}) U^\top$, where $P = U \operatorname{diag}_s(\sigma_s) U^\top$ is an eigendecomposition of P . We denote $\log : \mathcal{S}_{++}^d \rightarrow \mathcal{S}^d$ the matrix logarithm $\log(P) = U \operatorname{diag}_s(\log \sigma_s) U^\top$, which is the inverse of \exp on \mathcal{S}_{++}^d . We adopt some conventions in order to deal conveniently with singular matrices. We extend $(P, Q) \mapsto P \log(Q)$ by lower semicontinuity on $(\mathcal{S}_+^d)^2$, i.e., writing $Q = U \operatorname{diag}_s(\sigma_s) U^\top$ and $\tilde{P} = U^\top P U$,

$$P \log Q := \begin{cases} P \log Q & \text{if } \ker Q = \emptyset, \\ U[\tilde{P} \operatorname{diag}_s(\log \sigma_s)] U^\top & \text{if } \ker Q \subset \ker P, \\ +\infty & \text{otherwise,} \end{cases}$$

with the convention $0 \log 0 = 0$ when computing the matrix product in square brackets. Moreover, for $(P, (Q_i)_{i \in I}) \in \mathcal{S}^d \times (\mathcal{S}_+^d)^I$, the matrix $\exp(P + \sum_i \log Q_i)$ is by convention the matrix in \mathcal{S}_+^d which kernel is $\sum_i \ker Q_i$ (and is unambiguously defined on the orthogonal of this space).

¹ <https://github.com/gpeyre/2017-EJAM-quantum-ot>

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

$$H(\mu) \stackrel{\text{def}}{=} \sum_i H(\mu_i) \quad \text{where} \tag{1.1}$$

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

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 $\mu = (\mu_i)_i, \zeta = (\zeta_i)_i$ in \mathcal{S}_+^d corresponding to measures defined on the same grid, it is defined as

$$\text{KL}(\mu|\zeta) \stackrel{\text{def}}{=} \sum_i \text{KL}(\mu_i|\zeta_i), \tag{1.2}$$

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

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

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

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

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

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

2 Kantorovich problem for tensor-valued transport

We consider two measures that are sums of Dirac masses

$$\mu = \sum_{i \in I} \mu_i \delta_{x_i} \quad \text{and} \quad \nu = \sum_{j \in J} \nu_j \delta_{y_j} \tag{2.1}$$

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 [36], 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 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. [47], in contrast to the scalar case, in the matrix case (dimension $d > 1$), this constraint is in general 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 the unbalanced case, and following [41], 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 KulLback-Leibler divergence (which in particular gives rise to a natural Riemannian structure on positive measures, as defined in [21, 37, 42]). We propose to make use of its quantum counterpart (1.2) via the following convex program

$$W(\mu, \nu) = \min_{\gamma} \langle \gamma, c \rangle + \rho_1 \text{KL}(\gamma \mathbb{1}_J | \mu) + \rho_2 \text{KL}(\gamma^\top \mathbb{1}_I | \nu) \tag{2.2}$$

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 $d = 1$, (2.2) recovers exactly the log-entropic definition [41] of unbalanced OT, which is studied numerically by Chizat et al. [22]. 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 (2.2) 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 at the same location x and associated tensors $(P, Q) \in (\mathcal{S}_+^d)^2$, one obtains

the following “metric” between tensors (assuming $\rho_1 = \rho_2 = 1$ for simplicity)

$$\sqrt{W(P\delta_x, Q\delta_x)} = D(P, Q) \stackrel{\text{def.}}{=} \text{tr}(P + Q - 2\mathfrak{M}(P, Q))^{\frac{1}{2}} \tag{2.3}$$

where $\mathfrak{M}(P, Q) \stackrel{\text{def.}}{=} \exp(\log(P)/2 + \log(Q)/2)$. When (P, Q) commute, one has $D(P, Q) = \|\sqrt{P} - \sqrt{Q}\|$, which is a distance. In the general case, we do not know whether D is a distance (basic numerical tests do not exclude this property).

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 (2.2) 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 (away from singular points) following the method of Crane *et al.* [23]. In this setting, the cost is usually chosen to be $c_{i,j} = d_X(x_i, x_j)^2 \text{Id}_{d \times d}$, where d_X is the geodesic distance on X .

Remark 4 (Measure lifting). An alternative to compute OT between tensor fields would be to rather lift the input measure μ to a measures $\bar{\mu} \stackrel{\text{def.}}{=} \sum_{i \in I} \delta_{(\mu_i, x_i)}$ defined over the space $X \times \mathcal{S}_+^d$ (and similarly for the lifting $\bar{\nu}$ of ν) and then use traditional scalar OT over this lifted space (using a ground cost taking into account both space and tensor variations). Such a naive approach would destroy the geometry of tensor-valued measures, and result in very different interpolations. For example, a sum of two nearby Diracs on $X = \mathbb{R}$

$$\mu = P\delta_0 + Q\delta_s \quad \text{where} \quad P \stackrel{\text{def.}}{=} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Q \stackrel{\text{def.}}{=} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

is treated by our method as being very close to $\text{Id}_{2 \times 2} \delta_0$ (which is the correct behaviour of a *measure*), whereas it would be lifted to $\bar{\mu} = \delta_{(0,P)} + \delta_{(s,Q)}$ over $\mathbb{R} \times \mathcal{S}_+^2$, which is in contrast very far from $\delta_{(0, \text{Id}_{2 \times 2})}$.

2.2 Quantum transport interpolation

Given two input measures (μ, ν) , we denote by γ a solution of (2.2) or, in practice, its regularized version (see (3.1) 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 3.4 shows some numerical illustrations of this interpolation. Note also that Section 4 proposes a generalization of this idea to compute an interpolation (barycentre) between more than two input fields.

Mimicking the definition of the OT interpolation (the so-called McCann displacement interpolation; see for instance [52]), 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 $c_{i,j} = d_X(x_i, y_j)^2 \text{Id}_{d \times d}$ 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

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

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

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

the adjustment factors which account for the imperfect match of the marginal associated to a solution of (3.1); the adjusted coupling is

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

Finally, the interpolating measure is then defined as

$$\forall t \in [0, 1], \quad \mu_t \stackrel{\text{def.}}{=} \sum_{i,j} \gamma_{i,j}^t \delta_{x_{i,j}^t}. \tag{2.4}$$

One easily verifies that this measure indeed interpolates the two input measures, i.e., $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$. This formula (2.4) 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 $t = 1$).

Remark 5 (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 (2.4). Note that the entropic regularization algorithm detailed in Section 3 destroys this exact sparsity, but we found numerically that thresholding to zero the small entries of γ generates accurate approximations.

3 Quantum Sinkhorn

The convex program (2.2) 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 [24], we propose to use a similar entropic regularized strategy (see also Section 1), but this time with the quantum entropy (1.1).

3.1 Entropic regularization

We define an entropic regularized version of (2.2)

$$W_\varepsilon(\mu, \nu) \stackrel{\text{def.}}{=} \min_\gamma \langle \gamma, c \rangle + \rho_1 \operatorname{KL}(\gamma \mathbb{1}_J | \mu) + \rho_2 \operatorname{KL}(\gamma^\top \mathbb{1}_I | \nu) - \varepsilon H(\gamma). \tag{3.1}$$

Note that when $\varepsilon = 0$, one recovers the original problem (2.2). This is a strongly convex program, with a unique solution. The crux of this approach, as already known in the

scalar case (see [22]), 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 (3.1) reads*

$$W_\varepsilon(\mu, v) = \max_{u,v} -\text{tr} \left[\rho_1 \sum_i (e^{u_i + \log(\mu_i)} - \mu_i) + \rho_2 \sum_j (e^{v_j + \log(v_j)} - v_j) + \varepsilon \sum_{i,j} e^{\mathcal{K}(u,v)_{i,j}} \right], \quad (3.2)$$

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

$$\mathcal{K}(u, v)_{i,j} \stackrel{\text{def.}}{=} -\frac{c_{i,j} + \rho_1 u_i + \rho_2 v_j}{\varepsilon}. \quad (3.3)$$

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

$$\forall (i, j), \quad \gamma_{i,j} = \exp(\mathcal{K}(u, v)_{i,j}). \quad (3.4)$$

Proof Applying the Fenchel–Rockafellar duality theorem [50] to (3.1) leads to the dual program

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

where $\text{KL}^*(\cdot | \mu)$ 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–Fenchel transform 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 [29] (see [6] for its extension to Bregman divergences) to solve (3.2). This corresponds to alternatively maximizing (3.2) 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,

$$\text{LSE}_j(K) \stackrel{\text{def.}}{=} \left(\log \sum_j \exp(K_{i,j}) \right)_i, \quad (3.5)$$

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 (3.2) satisfies*

$$\forall i, \quad u_i = \text{LSE}_j(\mathcal{K}(u, v))_i - \log(\mu_i), \quad (3.6)$$

$$\forall j, \quad v_j = \text{LSE}_i(\mathcal{K}(u, v))_j - \log(v_j), \quad (3.7)$$

where $\mathcal{K}(u, v)$ is defined in (3.3).

Proof Writing the first order condition of (3.2) 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 the explicit alternating minimization with respect to u and v in Dykstra’s with just one step of fixed point iteration (3.6) and (3.7). 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 (3.6) and (3.7). We use the following τ -relaxed assignment notation

$$a \stackrel{\tau}{\leftarrow} b \quad \text{means that} \quad a \leftarrow (1 - \tau)a + \tau b. \tag{3.8}$$

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

Remark 6 (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 [22], and each update of u (resp. v) exactly solves the fixed point (3.6) (resp. (3.7)). 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. Figure 3 displays a typical example of convergence, and exemplifies the usefulness of using large values of τ_k , which leads to a speed-up of a factor 6 with respect to the usual Sinkhorn’s choice $\tau_k = \frac{\varepsilon}{\varepsilon + \rho_k}$.

Remark 7 (Heavy ball and acceleration). The extrapolation steps with weights (τ_1, τ_2) are reminiscent of the heavy ball method [49] to accelerate the convergence of first order optimization methods. Non-linear acceleration methods would also be applicable, and would lift the need to manually tune the parameters (τ_1, τ_2) , see for instance [5]. Similar acceleration technics can also be derived by using accelerated first order schemes (such as Nesterov’s algorithm [45] or FISTA [7]) directly on the dual problem (3.2). Using second order information through quasi-Newton (L-BFGS) or conjugate-gradient is another option. We leave for future work the exploration of these variants of our basic iterates.

Remark 8 (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.

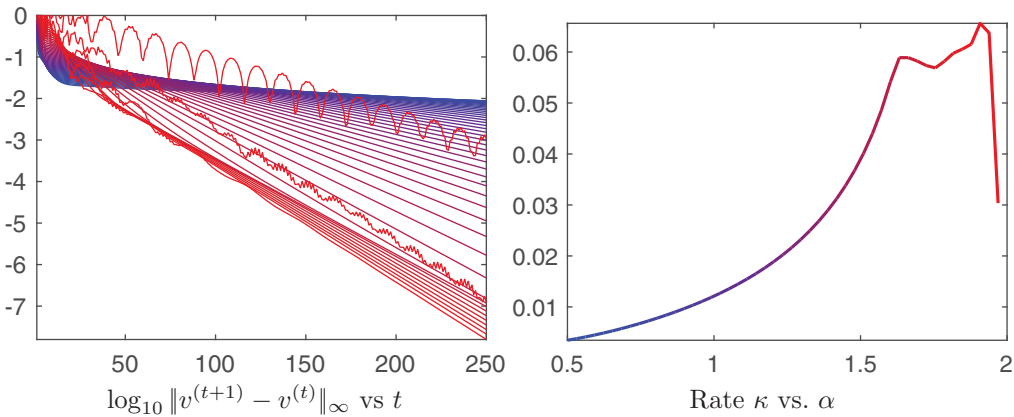


FIGURE 3. Display of convergence of Sinkhorn Algorithm 1 for the example displayed on the first row of Figure 1. Denoting $v^{(t)}$ the value of the variable v at iteration t , the left plot shows the fixed point residual error for increasing values of $\tau_1 = \tau_2 = \frac{\alpha \varepsilon}{\varepsilon + \rho_k}$ 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( $\mu, v, c, \varepsilon, \rho_1, \rho_2$ )
 $\forall k = 1, 2, \quad \tau_k \in ]0, \frac{2\varepsilon}{\varepsilon + \rho_k} [$ 
 $\forall (i, j) \in I \times J, \quad (u_i, v_j) \leftarrow (0_{d \times d}, 0_{d \times d})$ 
for  $s = 1, 2, 3, \dots$ 
     $K \leftarrow \mathcal{K}(u, v)$ 
     $\forall i \in I, \quad u_i \xleftarrow{\tau_1} \text{LSE}_j(K_{i,j}) - \log(\mu_i)$ 
     $K \leftarrow \mathcal{K}(u, v)$ 
     $\forall j \in J, \quad v_j \xleftarrow{\tau_2} \text{LSE}_i(K_{i,j}) - \log(v_j)$ 
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 (3.1). The operator \mathcal{K} is defined in (3.3).

Remark 9 (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 in Section 5 only require 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 [2, 3].

Remark 10 (Computational complexity). For low-dimensional problems (typically for those considered in Section 5), the Q-Sinkhorn Algorithm 1 scales to grid sizes of roughly 5k 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 solv-

ing (3.1) is highly sparse (up to machine precision for small enough ε). We thus found that the use of the multi-scale refinement strategy introduced in [53] is able to make the Q-Sinkhorn scale to high resolution grids. It is not used to produce the figures of this article, but it is available in the companion computational toolbox.

Remark 11 (Gurvits’ non-commutative Sinkhorn). Let us insist on the fact that the proposed Q-Sinkhorn Algorithm 1 is unrelated to Gurvits’ Sinkhorn algorithm [33]. While Gurvits’ iterations compute a coupling between a pair of input tensors, our method rather couples two *fields* of tensors (viewed as tensor-valued measures). Our usage of the wording “quantum” refers to the notion of quantum entropy (1.1) and is not inspired by quantum physics.

3.3 Trace-constrained extension

The quantum OT problem (2.2) does not impose that the marginals of the coupling γ match exactly the inputs (μ, ν) . It is only in the limit $(\rho_1, \rho_2) \rightarrow (+\infty, +\infty)$ that an exact match is obtained, but as explained in Section 2.1, this might leads to an empty constraint set.

To address this potential issue, we propose to rather only impose the *trace* of the marginals to match the trace of the input measures, in order to guarantee conservation of mass (as measured by the trace of the tensors). We thus propose to solve the entropy regularized problem (3.1) with the extra constraint

$$\forall i \in I, \quad \sum_j \text{tr}(\gamma_{i,j}) = \text{tr}(\mu_i) \quad \text{and} \quad \forall j \in J, \quad \sum_i \text{tr}(\gamma_{i,j}) = \text{tr}(\nu_j).$$

These two extra constraints introduce two dual Lagrange multipliers $(\alpha, \beta) \in \mathbb{R}^I \times \mathbb{R}^J$ and the optimal coupling relation (3.4) is replaced by

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

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

Q-Sinkhorn algorithm 1 is extended to handle these two extra variables (α, β) by simply adding two steps to update these variables

$$\begin{aligned} \forall i \in I, \quad \alpha_i &\leftarrow \alpha_i + \varepsilon \text{LSTE}_j(K)_i \quad \text{where } K \stackrel{\text{def.}}{=} \mathcal{K}(u, v, \alpha, \beta), \\ \forall j \in J, \quad \beta_j &\leftarrow \beta_j + \varepsilon \text{LSTE}_i(K)_j \quad \text{where } K \stackrel{\text{def.}}{=} \mathcal{K}(u, v, \alpha, \beta). \end{aligned}$$

where we introduced the log-sum-trace-exp operator

$$\text{LSTE}_j(K) \stackrel{\text{def.}}{=} \left(\log \sum_j \text{tr}(\exp(K_{i,j})) \right)_i$$

(and similarly for LSTE_i). Note that in this expression, the \exp is matrix-valued, whereas the \log is real-valued.

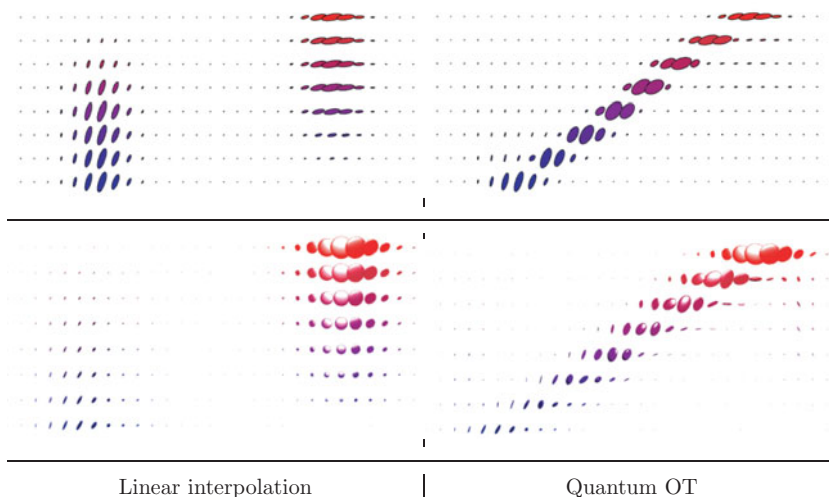


FIGURE 4. Comparison of linear and quantum-OT interpolation (using formula (2.4)). Each row shows a tensor field μ_t (top $d = 2$, bottom $d = 3$) along a linear segment from $t = 0$ to $t = 1$ (t axis is vertical).

3.4 Numerical illustrations

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

- (1) One first computes the optimal γ solving (3.1) using Sinkhorn iterations (Algorithm 1).
- (2) Then, for any $t \in [0, 1]$, one computes μ_t using this optimal γ with formula (2.4).

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 [52]. 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. This example illustrates that the behaviour of our tensor interpolation is radically different from only applying classical scalar-valued OT to the trace of the tensor (which would result in the same coupling as the one obtained with isotropic tensors, Figure 5, left).

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 $c_{i,j} = d_X(x_i, y_j)^2 \text{Id}_{2 \times 2}$.

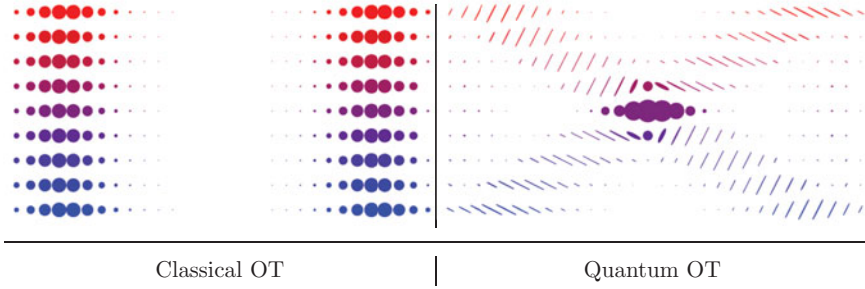


FIGURE 5. Comparison of classical OT (i.e., between isotropic tensors) and quantum-OT (between anisotropic tensors) interpolation (using formula (2.4)), using the same display as Figure 4.

4 Quantum barycentres

Following Agueh and Carlier [1] (see also [9, 56] for numerical methods using entropic regularization), we now propose a generalization of the OT problem (2.2), where, instead of coupling only two input measures, one tries to couple an arbitrary set of inputs, and compute their Fréchet means.

4.1 Barycentre optimization problem

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

$$\min_v \sum_\ell w_\ell W_\varepsilon(\mu^\ell, v), \tag{4.1}$$

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 (2.2) of W_ε . Note that the choice $\rho_2 = +\infty$ corresponds to imposing the exact hard marginal constraint $\gamma^\top \mathbb{1}_J = v$.

Remark 12 (Barycentres 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 barycentre (unregularized, i.e., $\varepsilon = 0$) for a cost $d_X(x, y)^q \text{Id}_{d \times d}$ is $P \delta_{x^*}$ where $x^* \in X$ is the usual barycentre for the distance d_X , solving

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

and the barycentric matrix is

$$P = e^{-\frac{\varepsilon(x^*)}{\rho}} \exp \left(\sum_\ell w_\ell \log(P_\ell) \right). \tag{4.2}$$

Figure 6 illustrates the effect of a pointwise interpolation (i.e., at the same location x_ℓ for all ℓ) between tensors.

Problem (4.1) is convex, and similarly to (3.2), it can be rewritten in dual form.

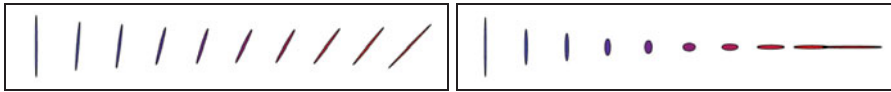


FIGURE 6. Two examples of pointwise (without transportation) interpolations, using formula (4.2). Here P_1 and P_2 are represented using the blue/red ellipses on the left/right, and weights are $(w_1, w_2) = (1 - t, t)$ for $t \in [0, 1]$ from left to right.

Proposition 3. *The optimal v solving (4.1) is the solution of*

$$\max_{(u^\ell, v^\ell)} \min_v - \sum_{\ell} w_{\ell} \operatorname{tr} \left[\rho \sum_i e^{u_i^\ell + \log(\mu_i^\ell)} + \sum_j v_j v_j^\ell + \varepsilon \sum_{i,j} e^{\mathcal{K}(u^\ell, v^\ell)_{i,j}} \right], \tag{4.3}$$

where here we define \mathcal{K} as

$$\mathcal{K}(u, v)_{i,j} \stackrel{\text{def.}}{=} - \frac{c_{i,j} + \rho u_i + v_j}{\varepsilon}. \tag{4.4}$$

4.2 Quantum barycentre Sinkhorn

Similarly to Proposition 2, the dual solutions of (4.3) satisfy a set of coupled fixed point equations:

Proposition 4. *Optimal $(u^\ell, v^\ell)_\ell$ for (4.3) satisfy*

$$\forall (i, \ell), \quad \operatorname{LSE}_j(\mathcal{K}(u^\ell, v^\ell)_{i,j}) - \log(\mu_i^\ell), = u_i^\ell \tag{4.5}$$

$$\forall (j, \ell), \quad \operatorname{LSE}_i(\mathcal{K}(u^\ell, v^\ell)_{i,j}) = \log(v_j) \tag{4.6}$$

$$\sum_{\ell} w_{\ell} v^\ell = 0. \tag{4.7}$$

Proof The proof of (4.5) and (4.6) is the same as the one of Proposition 2. Minimization of (4.3) on v leads to (4.7). □

The extension of the quantum Sinkhorn algorithm to solve the barycentre problem (2) is detailed in Algorithm 2. It alternates between the updates of u , v and v , using the relaxed version of the fixed point equations (4.5), (4.6) and (4.7). The notation $\stackrel{\tau}{\leftarrow}$ refers to a relaxed assignment as defined in (3.8).

Remark 13 (Choice of τ). Remark 6 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 in [22]. We found experimentally that this choice leads to contracting (and hence linearly converging) iterations, and that higher values of τ usually accelerate the convergence rate.

Remark 14 (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 barycentre algorithm in [22].

```

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

Algorithm 2: Quantum-Barycentre iterations to compute the optimal barycentre measure v solving (4.1). The operator \mathcal{K} is defined in (4.4).

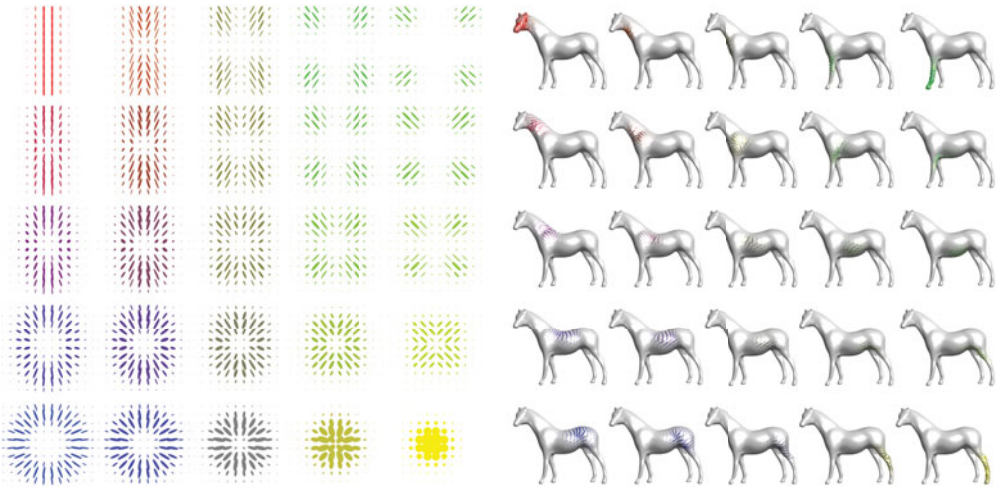


FIGURE 7. 5×5 barycentres of four input measures (displayed in the four corners). The weights $w \in \mathbb{R}^4$ correspond to bilinear interpolation weights (4.8) inside the square.

4.3 Numerical illustrations

Figure 7 shows examples of barycentres v solving (4.1) 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 grid) and parameterize the weights $(w_\ell)_{\ell=1}^4$ appearing in (4.1) as

$$(w_1, w_2, w_3, w_4) \stackrel{\text{def}}{=} ((1 - t_1)(1 - t_2), (1 - t_1)t_2, t_1(1 - t_2), t_1, t_2). \tag{4.8}$$

The left part of Figure 7 corresponds to measures on $X = Y = [0, 1]^2$ with $d = 2$ and ground cost $c_{i,j} = \|x_i - x_j\|^2 \text{Id}_{2 \times 2}$. The right part of Figure 7 corresponds to measures on

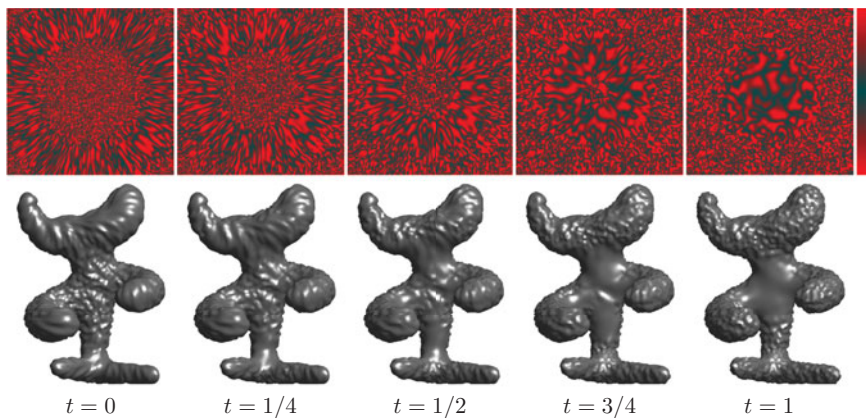


FIGURE 8. Example of interpolation between two input procedural anisotropic noise functions. The PSD tensor field parameterizing the texture are displayed on Figure 1. The colourmap used to render the anisotropic texture is displayed on the last column.

$X = Y$ 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

This section showcases four different applications of Q-OT to register and interpolate tensor fields. Unless otherwise stated, the data is normalized to the unit cube $[0, 1]^d$ (here $d = 2$ for images) and discretized on grids of $|I| = |J| = 50^d$ points. The regularization parameter is set to $\varepsilon = 0.08^2$, the fidelity penalty to $\rho = 1$, and the relaxation parameter for Sinkhorn to $\tau_k = \frac{1.8\varepsilon}{\varepsilon + \rho_k}$.

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 [39]. Following Lagae *et al.* [40] we consider here a spatially-varying Gabor noise function (i.e., non-stationary Gaussian noise), whose covariance function is parameterized using a PSD-valued field μ . Q-OT 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 [39] (which enables synthesis at arbitrary resolution and zooming). For simplicity, we rather consider here a more straightforward method, 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”.

Numerically, f is discretized on a 2-D grid, and μ is represented on this grid as a sum of Dirac masses (2.1). On Euclidean 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 colour texture by indexing a non-linear black-red colourmap (displayed on the right) using f_t . The 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 [4] and image compression [25]. 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 [54]). 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 (2.4)) 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 [11]. They corresponds to geodesic Delaunay triangulations for the Riemannian metric defined by the tensor field. This interpolation could typically be used to track the evolution of the solution of some PDE.

5.3 Diffusion tensor imaging

Diffusion tensor magnetic resonance imaging (DTI) is a popular technique to image the white matter of the brain (see [59] for a recent overview). DTI measures the diffusion of water molecules, which can be compactly encoded using a PSD tensor field $\mu(x) \in \mathcal{S}_+^3$, whose anisotropy and size matches the local diffusivity. A typical goal of this imaging technique is to map the brain anatomical connectivity, and in particular track the white matter fibres. This requires a careful handling of the tensor’s energy (its trace) and anisotropy, so using Q-OT is a perfect fit for such data.

Figure 10 shows an application of Q-OT for the interpolation (using 2.4) between 2-D slices from DTI tensor fields (μ, ν) acquired on two different subjects. This data is extracted from the studies [48, 57]. These two patients exhibit different anatomical connectivity geometries, and Q-OT is able to track the variation in both orientation and magnitude of the diffusion tensors. This figure also compares the different data fidelity parameters

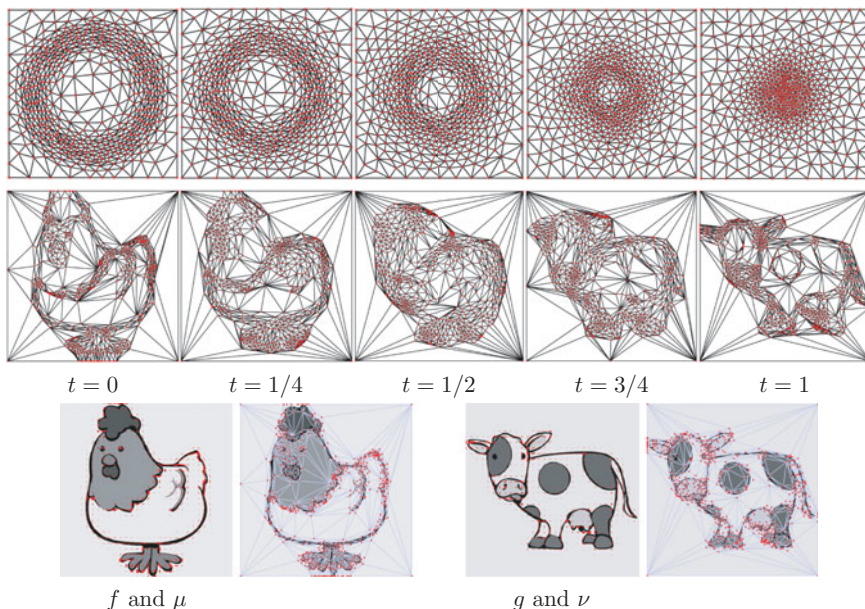


FIGURE 9. Two examples of interpolation between two input sizing fields $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$. *First row*: triangulation evolution for the sizing fields displayed on Figure 1. *Second row*: the input sizing fields $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$ are displayed on the third row,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

$\rho \in \{0.05, 1\}$. Selecting $\rho = 1$ enforces an overly-strong conservation constraint and leads to interpolation artifacts (in particular some structure are split during the interpolation). In contrast, selecting $\rho = 0.05$ introduces enough mass creation/destruction during the interpolation to be able to cope with strong inter-subject variability.

5.4 Spectral colour texture synthesis

As advocated initially in [31], a specific class of textured images (so-called micro-textures) is well-modelled using stationary Gaussian fields. In the following, we denote p the pixel positions and x the Fourier frequency indices. For colour images, these fields are fully characterized by their mean $m \in \mathbb{R}^3$ and their Fourier power spectrum, which is a tensor valued field $\mu(x)$ where, for each frequency x (defined on a 2-D grid) $\mu(x) \in \mathbb{C}^{3 \times 3}$ is a complex PSD hermitian matrix.

In practice, $\mu(x)$ is estimated from an exemplar colour image $f(p) \in \mathbb{R}^3$ using an empirical spectrogram

$$\mu(x) \stackrel{\text{def}}{=} \frac{1}{K} \sum_{k=1}^K \hat{f}_k(x) \hat{f}_k(x)^* \in \mathbb{C}^{3 \times 3}, \tag{5.1}$$

where \hat{f}_k is the Fourier transform of $f_k(p) \stackrel{\text{def}}{=} f(p)w_k(p)$ (computed using the FFT), w_k are windowing functions centered around K locations in the image plane, and $v^* \in \mathbb{C}^{1 \times 3}$ denotes the transpose-conjugate of a vector $v \in \mathbb{C}^{3 \times 1}$. Increasing the number K of

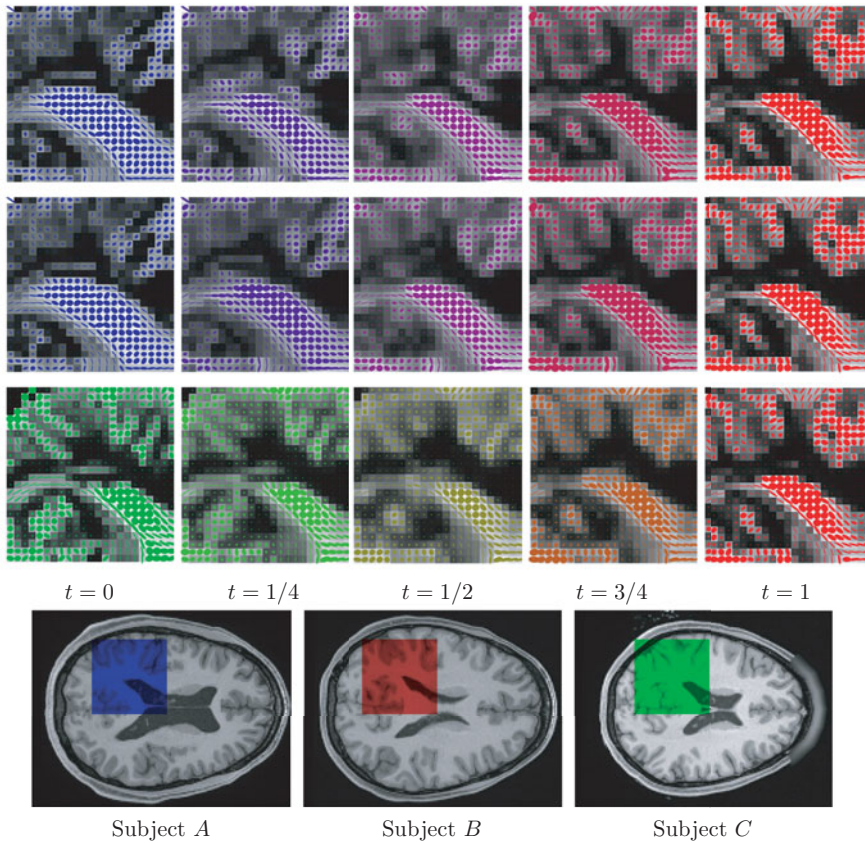


FIGURE 10. Interpolation between two 2-D slices of 3-D DTI tensor fields $(\mu, \nu) = (\mu_{t=0}, \mu_{t=1})$. For readability, only the X/Y components of the tensors are displayed. *First row*: interpolation between subjects (A, B) obtained using $\rho = 1$. *Second row*: interpolation between subjects (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.

windowed estimations helps to avoid having rank-deficient covariances ($K = 1$ leads to a field μ of rank-1 tensors).

Randomized new textures are then created by generating random samples $F(p) \in \mathbb{R}^3$ from the Gaussian field, which is achieved by defining the Fourier transform $\hat{F}(x) \stackrel{\text{def}}{=} m + \hat{N}(x)\sqrt{\mu(x)}\mathbb{1}_3$, where $N(p)$ is the realization of a Gaussian white noise, and $\sqrt{\cdot}$ is the matrix square root (see [31] for more details).

Figure 11 shows an application where two input power spectra (μ, ν) (computed using (5.1) from two input textures exemplars (f, g)) are interpolated using (2.4), and for each interpolation parameter $t \in [0, 1]$ a new texture F is synthesized and displayed. Note that while the Q-Sinkhorn Algorithm 1 is provided for real PSD matrices, it extends verbatim to complex positive Hermitian matrices (the matrix logarithm and exponential being defined the same way as for real matrices).

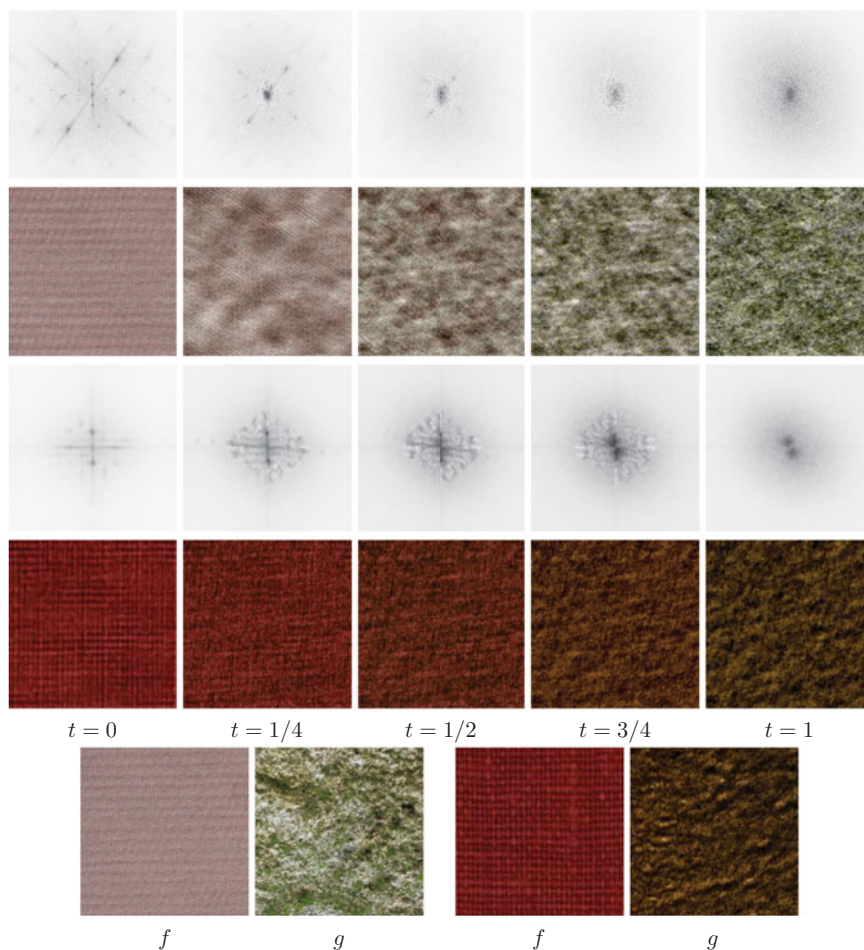


FIGURE 11. Row 1 and 3: display $\text{tr}(\mu_t(x))$ where μ_t are the interpolated power spectra. Rows 2 and 4: realizations of the Gaussian field parameterized by the power spectra μ_t . Row 5: input texture exemplars from which $(\mu_{t=0}, \mu_{t=1}) = (\mu, \nu)$ are computed.

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

The proposed formulation is quite versatile, and can be extended to other convex cones beyond PSD matrices, which is a promising direction for future work and applications. Other possible research directions also include investigating relationships between this static formulation of tensor-valued OT and dynamic formulations.

Acknowledgements

DTI data were provided by Franco Pestilli (NSF IIS 1636893; NIH ULTTR001108) and the Human Connectome Project (NIH 1U54MH091657). The images for the texture synthesis experiments are from IPOL.² The images for the triangular meshing experiments are from Wikimedia.

References

- [1] AGUEH, M. & CARLIER, G. (2011) Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43**(2), 904–924.
- [2] AL-MOHY, A. H. & HIGHAM, N. J. (2009) A new scaling and squaring algorithm for the matrix exponential. *SIAM J. Sci. Comput.* **31**(3), 970–989.
- [3] AL-MOHY, A. H. & HIGHAM, N. J. (2012) Improved inverse scaling and squaring algorithms for the matrix logarithm. *SIAM J. Sci. Comput.* **34**(4), C153–C169.
- [4] ALLIEZ, P., COHEN-STEINER, D., DEVILLERS, O., LÉVY, B. & DESBRUN, M. (2003) Anisotropic polygonal remeshing. In: *ACM Transactions on Graphics (TOG)*, Vol. 22, ACM, pp. 485–493.
- [5] ANDERSON, D. G. (1965) Iterative procedures for nonlinear integral equations. *J. ACM.* **12**(4), 547–560.
- [6] BAUSCHKE, H. H. & LEWIS, A. S. (2000) Dykstra’s algorithm with Bregman projections: a convergence proof. *Optimization.* **48**(4), 409–427.
- [7] BECK, A. & TBOULLE, M. (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202.
- [8] BENAMOU, J.-D. & BRENIER, Y. (2000) A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **84**(3), 375–393.
- [9] BENAMOU, J.-D., CARLIER, G., CUTURI, M., NENNA, L. & PEYRÉ, G. (2015) Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comp.* **37**(2), A1111–A1138.
- [10] BONNEEL, N., VAN DE PANNE, M., PARIS, S. & HEIDRICH, W. (2011) Displacement interpolation using Lagrangian mass transport. *ACM Trans. Graph.* **30**(6), 158:1–158:12.
- [11] BOUGLEUX, S., PEYRÉ, G. & COHEN, L. D. (2009) Image compression with anisotropic geodesic triangulations. In: *Proceedings of ICCV’09*, pp. 2343–2348.
- [12] 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. Math. Phys.* **7**(3), 200–217.
- [13] BRENIER, Y. (1991) Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44**(4), 375–417.
- [14] CARLEN, E. A. & 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. *Commun. Math. Phys.* **331**(3), 887–926.
- [15] CARLEN, E. A. & MAAS, J. (2017) Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *J. Funct. Anal.* **273**(5), 1810–1869.
- [16] CHANDRASEKARAN, V. & SHAH, P. (2017) Relative entropy optimization and its applications. *Math. Program.* **161**(1), 1–32.
- [17] CHEN, Y., GANGBO, W., GEORGIU, T. T. & TANNENBAUM, A. (2017) On the matrix Monge-Kantorovich problem. *Preprint arXiv:1701.02826*.
- [18] CHEN, Y., GEORGIU, T. T. & TANNENBAUM, A. (2016) Matrix optimal mass transport: A quantum mechanical approach. *Preprint arXiv:1610.03041*.
- [19] CHEN, Y., GEORGIU, T. T. & TANNENBAUM, A. (2016) Vector-valued optimal mass transport. *Preprint arXiv:1610.03041*.

² www.ipol.im

- [20] CHIZAT, L., PEYRÉ, G., SCHMITZER, B. & VIALARD, F.-X. (2015) Unbalanced optimal transport: Geometry and Kantorovich formulation. Preprint 1508.05216, Arxiv.
- [21] CHIZAT, L., PEYRÉ, G., SCHMITZER, B. & VIALARD, F.-X. (2016) An interpolating distance between optimal transport and Fisher–Rao metrics. *Found. Comput. Math.*, doi: 10.1007/s10208-016-9331-y.
- [22] CHIZAT, L., PEYRÉ, G., SCHMITZER, B. & VIALARD, F.-X. (2016) Scaling algorithms for unbalanced transport problems to appear in *Mathematics of Computation*, 2017. Preprint 1607.05816, Arxiv.
- [23] CRANE, K., DESBRUN, M. & SCHRÖDER, P. (2010) Trivial connections on discrete surfaces. In: *Computer Graphics Forum*, Vol. 29, Wiley Online Library, pp. 1525–1533.
- [24] CUTURI, M. (2013) Sinkhorn distances: Lightspeed computation of optimal transportation. In: *Proceedings of NIPS*, Vol. 26, pp. 2292–2300.
- [25] DEMARET, L., DYN, N. & ISKE, A. (2006) Image compression by linear splines over adaptive triangulations. *Signal Process.* **86**(7), 1604–1616.
- [26] DERICHE, R., TSCHUMPELÉ, D., LENGLET, C. & ROUSSON, M. (2006) *Variational Approaches to the Estimation, Regularization, and Segmentation of Diffusion Tensor Images*, Springer US, Boston, MA, pp. 517–530.
- [27] DHILLON, I. S. & TROPP, J. A. (2008) Matrix nearness problems with Bregman divergences. *SIAM J. Matrix Anal. Appl.*, **29**(4), 1120–1146.
- [28] DRYDEN, I. L., KOLOYDENKO, A. & ZHOU, D. (2009) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3**(3), 1102–1123.
- [29] DYKSTRA, R. L. (1983) An algorithm for restricted least squares regression. *J. Amer. Stat.* **78**(384), 839–842.
- [30] FROGNER, C., ZHANG, C., MOBAHI, H., ARAYA, M. & POGGIO, T. A. (2015) Learning with a Wasserstein loss. In: *Advances in Neural Information Processing Systems*, Vol. 28, pp. 2044–2052.
- [31] GALERNE, B., GOUSSEAU, Y. & MOREL, J.-M. (2011) Random phase textures: Theory and synthesis. *IEEE Trans. Image Process.* **20**(1), 257–267.
- [32] GEORGIU, T. T. & PAVON, M. (2015) Positive contraction mappings for classical and quantum Schrödinger systems. *J. Math. Phys.* **56**(3), 033301.
- [33] GURVITS, L. (2004) Classical complexity and quantum entanglement. *J. Comput. Syst. Sci.* **69**(3), 448–484.
- [34] HOTZ, I., FENG, L., HAGEN, H., HAMANN, B., JOY, K. & JEREMIC, B. (2004) Physically based methods for tensor field visualization. pages 123–130. IEEE Computer Society.
- [35] JIANG, X., NING, L. & GEORGIU, T. T. (2012) Distances and Riemannian metrics for multivariate spectral densities. *IEEE Trans. Autom. Control.* **57**(7), 1723–1735, Jul. 2012.
- [36] KANTOROVICH, L. V. (1942) On the transfer of masses (in Russian). *Dokl. Akad. Nauk* **37**(2), 227–229.
- [37] KONDRATYEV, S., MONSAINGEON, L. & VOROTNIKOV, D. (2016) A new optimal transport distance on the space of finite radon measures. *Adv. Differ. Equ.* **21**(11/12), 1117–1164, 11.
- [38] KULIS, B., SUSTIK, M. A. & DHILLON, I. S. (2009) Low-rank kernel learning with Bregman matrix divergences. *J. Mach. Learn. Res.* **10**, 341–376.
- [39] LAGAE, A., LEFEBVRE, S., COOK, R., DE ROSE, T., DRETTAKIS, G., EBERT, D. S., LEWIS, J. P., PERLIN, K. & ZWICKER, M. (2010) A survey of procedural noise functions. In *Computer Graphics Forum*, Vol. 29, Wiley Online Library, pp. 2579–2600.
- [40] LAGAE, A., LEFEBVRE, S. & DUTRÉ, P. (2011) Improving Gabor noise. *IEEE Trans. Vis. Comput. Graph.* **17**(8), 1096–1107.
- [41] LIERO, M., MIELKE, A. & SAVARÉ, G. (2015) Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. ArXiv e-prints.
- [42] LIERO, M., MIELKE, A. & SAVARÉ, G. (2016) Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves. *SIAM J. Math. Anal.* **48**(4), 2869–2911.

- [43] MITTENZWEIG, M. & MIELKE, A. (2017) An entropic gradient structure for lindblad equations and couplings of quantum systems to macroscopic models. *J. Stat. Phys.* **167**(2), 205–233.
- [44] MONGE, G. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*. 1781, pp. 666–704.
- [45] NESTEROV, Y. (1983) A method of solving a convex programming problem with convergence rate $o(1/k^2)$. In: *Soviet Mathematics Doklady*. 27, pp. 372–376.
- [46] NING, L. & GEORGIU, T. T. (2014) Metrics for matrix-valued measures via test functions. In: *Proceedings of the 53rd IEEE Conference on Decision and Control*, IEEE, pp. 2642–2647.
- [47] NING, L., GEORGIU, T. T. & TANNENBAUM, A. (2015) On matrix-valued Monge–Kantorovich optimal mass transport. *IEEE Trans. Autom. Control.* **60**(2), 373–382.
- [48] PESTILLI, F., YEATMAN, J. D., ROKEM, A., KAY, K. N. & WANDELL, B. A. (2014) Evaluation and statistical inference for human connectomes. *Nature Methods* **11**(10), 1058–1063.
- [49] POLYAK, B. T. (1964) Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **4**(5), 1–17.
- [50] ROCKAFELLAR, R.-T. (1970) *Convex Analysis*, Princeton University Press, Princeton, New Jersey.
- [51] RUBNER, Y., TOMASI, C. & GUIBAS, L. J. (2000) The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40**(2), 99–121.
- [52] SANTAMBROGIO, F. (2015) Optimal transport for applied mathematicians. In: Birkhäuser (editor), *Prog. Nonlinear Differ. Equ. Appl.* 87, Springer International Publishing AG Switzerland, NY.
- [53] SCHMITZER, B. (2016) Stabilized sparse scaling algorithms for entropy regularized transport problems. *arXiv:1610.06519*.
- [54] SHEWCHUK, J. (2002) *What is a good linear finite element? Interpolation, conditioning, anisotropy and quality measures* (preprint). 73, *University of California*, Berkeley.
- [55] SINKHORN, R. (1964) A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.* **35**, 876–879.
- [56] SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T. & GUIBAS, L. (2015) Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *TOG.* **34**(4), 66:1–66:11.
- [57] TAKEMURA, H., CAIAFA, C. F., WANDELL, B. A. & PESTILLI, F. (2016) Ensemble tractography. *PLoS Comput. Biol.* **12**(2), e1004692.
- [58] VAXMAN, A., CAMPEN, M., DIAMANTI, O., PANOZZO, D., BOMMES, D., HILDEBRANDT, K. & BEN-CHEN, M. (2016) Directional field synthesis, design, and processing. *Comput. Graph. Forum* **35**(2), 545–572.
- [59] WANDELL, B. A. (2016) Clarifying human white matter. *Annu. Rev. Neurosci.* **8**(39), 103–28.
- [60] WEICKERT, J. (1998) *Anisotropic Diffusion in Image Processing*, Vol. 1, Teubner Stuttgart, Stuttgart.