

Interpolation of matrices and matrix-valued densities: The unbalanced case[†]

YONGXIN CHEN¹, TRYPHON T. GEORGIU²
and ALLEN TANNENBAUM³

¹*Department of Electrical and Computer Engineering, Iowa State University, IA, USA
email: yongchen@iastate.edu*

²*Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA
email: tryphon@uci.edu*

³*Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY, USA
email: allen.tannenbaum@stonybrook.edu*

(Received 20 November 2017; revised 18 April 2018; accepted 18 April 2018)

We propose unbalanced versions of the quantum mechanical version of optimal mass transport that is based on the Lindblad equation describing open quantum systems. One of them is a natural interpolation framework between matrices and matrix-valued measures via a quantum mechanical formulation of Fisher-Rao information and the matricial Wasserstein distance, and the second is an interpolation between Wasserstein distance and Frobenius norm. We also give analogous results for the matrix-valued density measures, i.e., we add a spatial dependency on the density matrices. This might extend the applications of the framework to interpolating matrix-valued densities/images with unequal masses.

Key words: Optimal mass transport, quantum mechanics, matrix-valued densities, Fisher-Rao information, Wasserstein metric.

1 Introduction

Optimal mass transport (OMT), in addition to its intrinsic mathematical elegance, has proven to be a very powerful methodology for numerous problems in econometrics, systems and control, information theory, statistical filtering and estimation, computer vision, and signal/image processing [19, 20]. However, the mass preservation requirement of OMT is many times unnatural for real-world problems. For example, in image registration or optical flow, one must impose *ad hoc* normalizations on the imagery that do not have a physical justification. Accordingly, there have been a number of approaches based on the interpolation of the Wasserstein metric from OMT with some other metric such as L^2 [1, 2] or some information-theoretic distance [9, 13–15]. The seminal work of Benamou

[†] This project was supported by AFOSR grants (FA9550-15-1-0045 and FA9550-17-1-0435), ARO grant (W911NF-17-1-049), grants from the National Center for Research Resources (P41-RR-013218) and the National Institute of Biomedical Imaging and Bioengineering (P41-EB-015902), National Science Foundation (NSF ECCS-1509387), NCI grant (1U24CA18092401A1), NIA grant (R01 AG053991), Breast Cancer Research Foundation, and a grant from the National Institutes of Health (P30-CA-008748).

and Brenier [2] makes this possible via certain modifications of the underlying energy functional and corresponding continuity equation.

In the previous work, Chen *et al.* [5,6] developed a quantum mechanical framework for defining a Wasserstein distance on density matrices (positive semi-definite matrices with trace 1), via a variational framework with a continuity equation constraint, generalizing the work of [2]. Moreover, in [6], the matricial framework was extended to the matrix-valued density measures allowing one to add spatial dimensions to the density matrices in the theory. We should note that independently and at about the same time, similar approaches to that of [6] were formulated and analysed in [4,16]. A remarkable result in [4,16] is that the Lindblad equation is in fact the gradient flow of the quantum entropy with respect to the Wasserstein metric defined there, which resembles to the classical results in the scalar setting [11]. This result may have a impact on physics and quantum mechanics. In contrast, our work in [6] was motivated by engineering applications like diffusion tensor imaging, multivariate spectral analysis where the matrix-valued densities appear. One essential problem is the numerical implementation. Therefore, our goal was to develop a *computable* Wasserstein metric for these types of applications. An efficient algorithm is developed in [8].

We show in the present note, that the methodology and definitions in [6] of matrix analogues of the gradient and divergence, allow us to formulate in a rather straightforward manner a natural energy functional and continuity equation that generalize the information-theoretic unbalanced approaches that give interpolations of Wasserstein and Fisher-Rao [9,13] and Hellinger [15]. Similarly, the interpolation [1,2] of Wasserstein and L^2 allows a natural generalization. Both interpolations in the scalar setting have been studied and proven to be effective in theory and applications [1,2,9,13,15]. Thus, we derive unbalanced versions of OMT in the matrix case as well as for matrix-valued density measures. Further, these ideas lead to Riemannian-type metrics on positive definite matrices for which we derive results on gradient flows relative to certain energy functions. Finally, we show that the unbalanced problem may be formulated as one of the convex optimization just like the balanced cases in [6], which makes it applicable to a variety of practical problems.

The remainder of this paper may be summarized as follows. First, we sketch several versions of unbalanced mass transport for scalar measures in Section 2. The preliminaries on non-commutative optimal transport are discussed in Section 3. The main results on non-commutative unbalanced transport are provided in Section 4 for positive matrices. In Section 5, analogous results are given for matrix-valued positive measures in the unbalanced setting. Several numerical examples are provided in Section 6 to illustrate our framework, and finally conclusions and future directions are sketched in Section 7.

2 Background on unbalanced mass transport

In this section, we briefly introduce the basis of OMT and review two possible methods for an unbalanced version of OMT in the scalar case following [1,9,13]. The original formulation of OMT is

$$\inf_T \left\{ \int_{\mathbb{R}^m} c(x, T(x)) \rho_0(x) dx \mid T_{\#} \rho_0 = \rho_1 \right\}, \quad (2.1)$$

where $c(x, y)$ denotes the cost of moving unit mass from x to y , and ρ_0, ρ_1 are two given probability densities in $\mathbb{R}^m$. The operator $(\cdot)_\#$ represents push forward of measures. By relaxing the map T to a coupling π [12], we obtain the Kantorovich form

$$\inf_{\pi \in \Pi(\rho_0, \rho_1)} \int_{\mathbb{R}^m} c(x, y) \pi(dx, dy), \quad (2.2)$$

with $\Pi(\rho_0, \rho_1)$ denoting the set of all joint distributions (couplings) between ρ_0 and ρ_1 . When the cost function $c(x, y) = \|x - y\|^2$, the problem has extremely rich structures [22]. As first pointed out in [2], the OMT problem has the fluid dynamic formulation

$$W_2(\rho_0, \rho_1)^2 := \inf_{\rho, v} \int_0^1 \int_{\mathbb{R}^m} \rho(t, x) \|v(t, x)\|^2 dx dt \quad (2.3a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad (2.3b)$$

$$\rho(0, \cdot) = \rho_0(\cdot), \quad \rho(1, \cdot) = \rho_1(\cdot). \quad (2.3c)$$

The minimization is taken over all the smooth velocity fields v and density flows ρ . They only need to satisfy the continuity equation (2.3b) in the weak sense. The reader is refer to [22, Chapter 8] for more technical details. The optimal value gives the (squared) Wasserstein distance W_2 . We next sketch two possible ways of extending W_2 to the unbalanced measures. These formulations are used to study unbalanced transport problems for the matrices and the matrix-valued densities.

2.1 L^2 and OMT

As noted in [1, 2], the L^2 distance can be used in conjunction with OMT in case of unbalanced mass distributions. The dynamic framework of [2] makes this quite straightforward. Full details and numerics may be found in [1].

Accordingly, given two unbalanced densities ρ_0 and ρ_1 , it is natural to seek a distribution $\tilde{\rho}_1$ close to ρ_1 in the L^2 sense, which keeps the Wasserstein distance $W_2(\rho_0, \tilde{\rho}_1)^2$ small in the mean time. The L^2 perturbation may be interpreted as ‘noise’. One can then show that this problem amounts to minimizing

$$\inf_{\rho, v, \tilde{\rho}_1} \int_0^1 \int_{\mathbb{R}^m} \rho(t, x) \|v\|^2 dx dt + \alpha \int_{\mathbb{R}^m} (\rho_1(x) - \tilde{\rho}_1(x))^2 dx, \quad (2.4a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad (2.4b)$$

$$\rho(0, \cdot) = \rho_0(\cdot), \quad \rho(1, \cdot) = \tilde{\rho}_1(\cdot), \quad (2.4c)$$

with $\alpha > 0$ being the coefficient balancing the two parts of the cost. This method has been used in several applications including optical flow, see [17] and the references therein.

Below is a slightly different form of interpolation distance between W_2 and L^2 that allows Riemannian-type structure. We bring in a source term s in the continuity equation

and construct a convex optimization problem

$$\inf_{\rho, v, s} \int_0^1 \int_{\mathbb{R}^m} \{ \rho(t, x) \|v\|^2 + \alpha s(t, x)^2 \} dx dt, \quad (2.5a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = s, \quad (2.5b)$$

$$\rho(0, \cdot) = \rho_0(\cdot), \quad \rho(1, \cdot) = \rho_1(\cdot). \quad (2.5c)$$

It can be shown that the square root of the minimum of the above is a well-defined metric on the space of probability densities with the finite second-order moments.

2.2 Unbalanced mass transport: Information theoretic formulations

We now review how Fisher-Rao may be employed to get an unbalanced formulation of OMT [9, 13, 15]. Unlike the method of [1, 2] where one interpolates L^2 and OMT, the idea here is to interpolate OMT and the Fisher-Rao metric. This is quite powerful since one explicitly combines an information-theoretic method with Wasserstein, and thus makes contact with the recent work on Schrödinger bridges [7]. There is a related method in [15] that interpolates between OMT and the Hellinger metric.

The interpolation of Fisher-Rao and Wasserstein is given as

$$\inf_{\rho, v, r} \int_0^1 \int_{\mathbb{R}^m} \{ \rho(t, x) \|v\|^2 + \alpha \rho(t, x) r^2 \} dx dt, \quad (2.6a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = \rho r, \quad (2.6b)$$

$$\rho(0, \cdot) = \rho_0(\cdot), \quad \rho(1, \cdot) = \rho_1(\cdot). \quad (2.6c)$$

Here, the minimum is taken over all the density flow ρ , the velocity field v , and the relative source intensity r satisfying the continuity equation with the source term (2.6b). Observing the source terms in (2.5b) and (2.6b) have the relation $s = \rho r$, we rewrite the second term of the cost in (2.6a) as

$$\int_0^1 \int_{\mathbb{R}^m} \rho(t, x) r(t, x)^2 dx dt = \int_0^1 \int_{\mathbb{R}^m} \frac{s(t, x)^2}{\rho(t, x)} dx dt. \quad (2.7)$$

This should be compared to

$$\int_0^1 \int_{\mathbb{R}^m} s(t, x)^2 dx dt, \quad (2.8)$$

which is used in (2.5). The cost (2.8) corresponds to the L^2 metric, while (2.7) defines the Fisher-Rao between the two smooth densities as

$$d_{\text{FR}}(\rho_0, \rho_1)^2 := \inf_{\rho, s} \int_0^1 \int_{\mathbb{R}^m} \frac{s(t, x)^2}{\rho(t, x)} dx dt, \\ \frac{\partial \rho}{\partial t} = s, \\ \rho(0, \cdot) = \rho_0(\cdot), \quad \rho(1, \cdot) = \rho_1(\cdot).$$

3 Quantum continuity equation

We sketch here the necessary background from [6], where full details can be found. Consider the two positive definite (Hermitian) matrices ρ_0 and ρ_1 . We seek a suitable generalization of the continuity equation that links the two matrices with a smooth path within the cone of positive matrices in suitable ways. In the context of quantum mechanics, ρ may represent a density matrix. In this, a standing assumption is that $\text{tr}(\rho_0) = \text{tr}(\rho_1) = 1$, and thereby, we seek paths $\rho(t)$ ($t \in [0, 1]$) between the two that maintain the same value for the trace.

Let $\mathcal{H}$, $\mathcal{S}$, and $\mathcal{C}$ denote the set of $n \times n$ Hermitian, skew-Hermitian matrices, and general complex matrices, respectively. Since matrices are $n \times n$ throughout, we dispense of n in the notation. We also denote the space of block-column vectors consisting of N elements in $\mathcal{H}$, $\mathcal{S}$, and $\mathcal{C}$ as $\mathcal{H}^N$, $\mathcal{S}^N$, and $\mathcal{C}^N$, respectively. Let now $\mathcal{H}_+$ and $\mathcal{H}_{++}$ denote the cones of non-negative and positive definite matrices, respectively, and

$$\mathcal{D} := \{\rho \in \mathcal{H}_+ \mid \text{tr}(\rho) = 1\},$$

$$\mathcal{D}_+ := \{\rho \in \mathcal{H}_{++} \mid \text{tr}(\rho) = 1\}.$$

Clearly, the tangent space of $\mathcal{D}_+$, at any $\rho \in \mathcal{D}_+$, is

$$T_\rho = T := \{\sigma \in \mathcal{H} \mid \text{tr}(\sigma) = 0\}.$$

We also use the standard notion of inner product

$$\langle X, Y \rangle = \text{tr}(X^* Y)$$

for both $\mathcal{H}$, $\mathcal{S}$, and $\mathcal{C}$. Here, $*$ denotes the complex transpose operator. For $X, Y \in \mathcal{H}^N$ ($\mathcal{S}^N, \mathcal{C}^N$),

$$\langle X, Y \rangle = \sum_{k=1}^N \text{tr}(X_k^* Y_k).$$

Given $X = [X_1^*, \dots, X_N^*]^* \in \mathcal{H}^N$ ($\mathcal{S}^N, \mathcal{C}^N$), $Y \in \mathcal{H}$ ($\mathcal{S}, \mathcal{C}$), denote

$$XY = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} Y := \begin{bmatrix} X_1 Y \\ \vdots \\ X_N Y \end{bmatrix},$$

$$YX = Y \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} := \begin{bmatrix} Y X_1 \\ \vdots \\ Y X_N \end{bmatrix},$$

and

$$\bar{X} = \begin{bmatrix} X_1^* \\ \vdots \\ X_N^* \end{bmatrix}.$$

In an open quantum system, the dynamics of the density matrices may be described by the Lindblad equation [10]

$$\dot{\rho} = -i(H\rho - \rho H) + \sum_{k=1}^N (L_k \rho L_k^* - \frac{1}{2} \rho L_k^* L_k - \frac{1}{2} L_k^* L_k \rho). \quad (3.1)$$

Here, the first term on the right-hand side describes the evolution of the state under the effect of the Hamiltonian H and it is energy preserving. The rest of the terms on the right-hand side represent the diffusion and capture the dissipation of energy. Note that this is the quantum analogue of the Laplacian operator Δ .

In the following, assume $L_k = L_k^*$, i.e., $L_k \in \mathcal{H}$ for all $k \in 1 \dots, N$. Under this assumption, we can define

$$\nabla_L : \mathcal{H} \rightarrow \mathcal{S}^N, \quad X \mapsto \begin{bmatrix} L_1 X - X L_1 \\ \vdots \\ L_N X - X L_N \end{bmatrix} \quad (3.2)$$

as the gradient operator. The dual of ∇_L , which is an analogue of the divergence operator, is given by

$$\nabla_L^* : \mathcal{S}^N \rightarrow \mathcal{H}, \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \mapsto \sum_k^N L_k Y_k - Y_k L_k. \quad (3.3)$$

This follows directly from the definition

$$\langle \nabla_L X, Y \rangle = \langle X, \nabla_L^* Y \rangle.$$

With this, we define the *Laplacian* as

$$\Delta_L X = -\nabla_L^* \nabla_L X = \sum_{k=1}^N (2L_k \rho L_k^* - \rho L_k^* L_k - L_k^* L_k \rho),$$

which is exactly (with some scaling) the diffusion term in the Lindblad equation (3.1). Therefore, the Lindblad equation (under the assumption $L_k = L_k^*$) can be rewritten as

$$\dot{\rho} = -i(H\rho - \rho H) + \frac{1}{2} \Delta_L \rho.$$

Note that the gradient operator ∇_L acts just like the standard gradient operator. Note that, in particular,

$$\nabla_L(XY + YX) = \nabla_L X Y + X \nabla_L Y + \nabla_L Y X + Y \nabla_L X, \quad \forall X, Y \in \mathcal{H}.$$

Using this gradient operator (3.2), we can then come up with several notions of the continuity equation [6]. In the present note for the interpolation of Fisher-Rao and Wasserstein, we will use

$$\dot{\rho} = \frac{1}{2} \nabla_L^*(v\rho + \rho v), \quad (3.4)$$

where the ‘velocity’ field $v = [v_1^*, \dots, v_N^*]^* \in \mathcal{S}^N$. Note $v\rho + \rho v \in \mathcal{S}^N$, which is consistent with the definition of ∇_L^* .

Usually, in the Lindblad equation (3.1), N is taken to be $n^2 - 1$. However, in general, we may choose $N \leq n^2 - 1$, as needed, possibly large enough such that in (3.4) we are able to cover the whole tangent space T_ρ at ρ for all $\rho \in \mathcal{D}_+$. In particular, we need ∇_L to have the property that the identity matrix I spans its null space. For instance, one can choose $L_1, \dots, L_N$ to be a basis of the Hermitian matrices $\mathcal{H}$, in which case $N = n(n+1)/2$. A simpler choice of basis L is

$$L_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \quad L_2 = \text{diag}([1, 2, \dots, n-1, 0]).$$

Obviously, both constructions ensure that the null space of ∇_L is spanned by I .

Now we have everything ready to define the fluid dynamic formulation of optimal transport on the space $\mathcal{D}_+$ of density matrices generalizing [2]. Given two density matrices $\rho_0, \rho_1 \in \mathcal{D}_+$, we formulate the following optimization problem:

$$W_2(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{D}_+, v \in \mathcal{S}^N} \int_0^1 \text{tr}(\rho v^* v) dt \quad (3.5a)$$

$$\dot{\rho} = \frac{1}{2} \nabla_L^*(v\rho + \rho v), \quad (3.5b)$$

$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1, \quad (3.5c)$$

where the minimum is taken over all the continuously differentiable density flows $\rho(\cdot) \in C^1([0, 1]; \mathcal{D}_+)$ and continuous velocity fields $v(\cdot) \in C([0, 1]; \mathcal{S}^N)$. The ‘Wasserstein distance’ between ρ_0 and ρ_1 is defined to be the square root of the minimum of the cost (3.5a). Note here for $v \in \mathcal{S}^N$, $v^*v = \sum_{k=1}^N v_k^* v_k$.

The Wasserstein distance function $W_2(\rho, \rho + \delta)$ gives an Riemannian-type structure on the tangent space T_ρ . More precisely, given two tangent vectors $\delta_1, \delta_2 \in T_\rho$, the Riemannian-type metric is given by

$$\langle \delta_1, \delta_2 \rangle_\rho = \frac{1}{2} \text{tr}(\rho(\nabla_L \lambda_1)^* \nabla_L \lambda_2 + \rho(\nabla_L \lambda_2)^* \nabla_L \lambda_1), \quad (3.6)$$

where λ_1 and λ_2 are the unique (up to an addition by a constant multiplication of I) solutions [5] to

$$\delta_j = \frac{1}{2} \nabla_L^*(\nabla_L \lambda_j \rho + \rho \nabla_L \lambda_j), \quad j = 1, 2.$$

In fact, $v = \nabla_L \lambda_j$ is the unique minimizer [5] of $\text{tr}(\rho v^* v)$ subject to the constraint

$$\delta_j = \frac{1}{2} \nabla_L^*(v\rho + \rho v).$$

Proposition 1

$$W_2(\rho_0, \rho_1) = \inf_{\rho \in \mathcal{D}_+} \int_0^1 \sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}} dt, \quad (3.7)$$

where the minimum is taken over all the continuously differentiable paths on the manifold $\mathcal{D}_+$ satisfying the boundary conditions $\rho(0) = \rho_0, \rho(1) = \rho_1$.

Proof Apparently,

$$W_2(\rho_0, \rho_1) \geq \inf_{\rho \in \mathcal{D}_+} \int_0^1 \sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}} dt.$$

To see this, note that for any $\rho(\cdot) \in C^1([0, 1]; \mathcal{D}_+)$, we have

$$\begin{aligned} \int_0^1 \sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}} dt &\leq \sqrt{\int_0^1 \langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)} dt} \\ &= \sqrt{\int_0^1 \text{tr}(\rho(\nabla_L \lambda)^* \nabla_L \lambda) dt}, \end{aligned}$$

where λ satisfies $\dot{\rho} = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda)$. The rest follows.

Now let us turn to the other direction. Let $\rho(\cdot) \in C^1([0, 1]; \mathcal{D}_+)$ be a (almost) minimizer of (3.7) with velocity field $\nabla_L \lambda$. Denote

$$\ell(t) = \int_0^t \sqrt{\langle \dot{\rho}(\tau), \dot{\rho}(\tau) \rangle_{\rho(\tau)}} d\tau,$$

and $s(t) = \ell(t)/\ell(1)$. Using a standard reparameterization

$$\hat{\rho}(s(t)) = \rho(t),$$

we obtain

$$\frac{d\hat{\rho}}{ds} = \frac{d\rho}{dt} \frac{dt}{ds} = \frac{1}{2} \nabla_L^* (\nabla_L \lambda(t) \rho(t) + \rho(t) \nabla_L \lambda(t)) \frac{\ell(1)}{\sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}}}.$$

Therefore, $\hat{\rho}(\cdot) \in C^1([0, 1]; \mathcal{D}_+)$, together with velocity field

$$\hat{v}(s(t)) = \nabla_L \hat{\lambda}(s(t)) := \nabla_L \lambda(t) \frac{\ell(1)}{\sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}}}$$

is also a feasible solution to (3.5). Now observing that

$$\text{tr}(\hat{\rho}(s) \hat{v}(s)^* \hat{v}(s)) = \text{tr}(\rho(t) (\nabla_L \lambda(t))^* \nabla_L \lambda(t)) \frac{\ell(1)^2}{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}} \equiv \ell(1)^2,$$

we conclude that $\int_0^1 \text{tr}(\hat{\rho}(s) \hat{v}(s)^* \hat{v}(s)) ds = \ell(1)^2$. Hence, $W_2(\rho_0, \rho_1)$ is upper bounded by $\ell(1)$, which completes the proof. $\square$

Therefore, $W_2(\cdot, \cdot)$ indeed defines a metric on $\mathcal{D}_+$ and it is intrinsic with respect to the Riemannian-type metric (3.6). If the condition that $\rho \in \mathcal{D}_+$ is relaxed to $\rho \in \mathcal{D}$, then an optimizer exists [5]. In general, it is difficult to establish the existence of the minimizer of (3.5) due to the possibility of ρ hitting the boundary of $\mathcal{D}$. One can, however, introduce a Lagrangian multiplier $\lambda(\cdot) \in C^1([0, 1]; \mathcal{H})$ for the constraints (3.5b), and arrive at the following sufficient conditions for optimality [6]. The equations (3.8) are also the geodesic equations in $\mathcal{D}_+$ with Riemannian-type metric (3.6).

Theorem 1 Suppose there exists $\lambda(\cdot) \in C^1([0, 1]; \mathcal{H})$ satisfying

$$\dot{\lambda} + \frac{1}{2}(\nabla_L \lambda)^*(\nabla_L \lambda) = \dot{\lambda} + \frac{1}{2} \sum_{k=1}^N (\nabla_L \lambda)_k^* (\nabla_L \lambda)_k = 0, \quad (3.8a)$$

and $\rho(\cdot) \in C^1([0, 1]; \mathcal{D}_+)$ satisfying

$$\dot{\rho} = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda), \quad (3.8b)$$

and matching the two marginals $\rho(0) = \rho_0, \rho(1) = \rho_1$, then $(\rho, v) = (\rho, \nabla_L \lambda)$ solves (3.5).

4 Interpolation of positive matrices

In this section, we describe the main results of the present note, namely the interpolation between quantum Wasserstein and Fisher-Rao and that between quantum Wasserstein and Frobenius norm, as the generalizations of (2.6) and (2.5), respectively. These provide ways to compare and interpolate positive matrices with possibly different traces.

4.1 Interpolation between Wasserstein and Fisher-Rao

Given $\rho_0, \rho_1 \in \mathcal{H}_{++}$ and $\alpha > 0$, define

$$W_{2,FR}(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{H}_{++}, v \in \mathcal{S}^N, r \in \mathcal{H}} \int_0^1 \{\text{tr}(\rho v^* v) + \alpha \text{tr}(\rho r^2)\} dt \quad (4.1a)$$

$$\dot{\rho} = \frac{1}{2} \nabla_L^* (v \rho + \rho v) + \frac{1}{2} (r \rho + \rho r), \quad (4.1b)$$

$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1. \quad (4.1c)$$

The minimum is taken over all the density flows $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ and velocity fields $v(\cdot) \in C([0, 1]; \mathcal{S}^N), r(\cdot) \in C([0, 1]; \mathcal{H})$. Note here the ‘continuity’ equation (4.1b), as a non-commutative generalization of (2.6). This distance $W_{2,FR}$ is an interpolation of W_2 and the Fisher-Rao distance

$$\begin{aligned} d_{FR}(\rho_0, \rho_1)^2 &:= \inf_{\rho \in \mathcal{H}_{++}, r \in \mathcal{H}} \int_0^1 \text{tr}(\rho r^2) dt, \\ \dot{\rho} &= \frac{1}{2} (r \rho + \rho r), \\ \rho(0) &= \rho_0, \quad \rho(1) = \rho_1, \end{aligned}$$

where the minimum is taken over $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ and $r(\cdot) \in C([0, 1]; \mathcal{H})$. Recall that the Bures metric [21] on the space of density matrices $\mathcal{D}_+$ is defined as

$$d_B(\rho, \rho + \delta)^2 = \frac{1}{2} \operatorname{tr}(G\delta),$$

where $G \in \mathcal{H}$ is the unique solution of

$$G\rho + \rho G = \delta$$

on $\mathcal{H}$. It follows

$$\frac{1}{2} \operatorname{tr}(G\delta) = \operatorname{tr}(\rho G^2).$$

Hence, d_{FR} is equivalent to the Bures metric d_B when restricted to space of density matrices $\mathcal{D}_+$. The distance $W_{2,\text{FR}}$ defines a Riemannian-type structure on $\mathcal{H}_{++}$.

Lemma 2 *For a given pair $\rho \in \mathcal{H}_{++}, \delta \in \mathcal{H}$, there is a unique solution to*

$$\delta = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda) + \frac{1}{2\alpha} (\lambda \rho + \rho \lambda).$$

Moreover, $(v, r) = (\nabla_L \lambda, \frac{1}{\alpha} \lambda)$ is the unique minimizer of

$$\operatorname{tr}(\rho v^* v) + \alpha \operatorname{tr}(\rho r^2)$$

over all the $(v \in \mathcal{S}^N, r \in \mathcal{H})$ satisfying

$$\delta = \frac{1}{2} \nabla_L^* (v\rho + \rho v) + \frac{1}{2} (r\rho + \rho r).$$

Proof Consider the convex optimization problem

$$\begin{aligned} \min_{v \in \mathcal{S}^N, r \in \mathcal{H}} \quad & \frac{1}{2} \operatorname{tr}(\rho v^* v) + \frac{\alpha}{2} \operatorname{tr}(\rho r^2) \\ \delta = \quad & \frac{1}{2} \nabla_L^* (v\rho + \rho v) + \frac{1}{2} (r\rho + \rho r). \end{aligned}$$

Clearly, it is feasible, so a minimizer exists. In addition, the cost is strictly convex, therefore the minimizer is unique. Using Karush–Kuhn–Tucker conditions [3], we conclude that the minimizer $(v_{\text{opt}}, r_{\text{opt}})$ minimizes the Lagrangian

$$\frac{1}{2} \operatorname{tr}(\rho v^* v) + \frac{\alpha}{2} \operatorname{tr}(\rho r^2) + \langle \lambda, \delta - \frac{1}{2} \nabla_L^* (v\rho + \rho v) - \frac{1}{2} (r\rho + \rho r) \rangle$$

for some multiplier $\lambda \in \mathcal{H}$. It is then straightforward to see $(v_{\text{opt}}, r_{\text{opt}}) = (\nabla_L \lambda, \frac{1}{\alpha} \lambda)$. The uniqueness of λ follows from that of the minimizer. This completes the proof. $\square$

Given the two tangent vectors $\delta_1, \delta_2 \in \mathcal{H}$ at ρ , the Riemannian-type metric is

$$\langle \delta_1, \delta_2 \rangle_\rho^{\text{FR}} = \frac{1}{2} \operatorname{tr}(\rho (\nabla_L \lambda_1)^* \nabla_L \lambda_2 + \rho (\nabla_L \lambda_2)^* \nabla_L \lambda_1) + \frac{1}{2\alpha} \operatorname{tr}(\rho \lambda_1 \lambda_2 + \rho \lambda_2 \lambda_1), \quad (4.2)$$

where $\lambda_i \in \mathcal{H}$, ($i = 1, 2$) is the unique solution to

$$\delta_i = \frac{1}{2} \nabla_L^* (\nabla_L \lambda_i \rho + \rho \nabla_L \lambda_i) + \frac{1}{2\alpha} (\lambda_i \rho + \rho \lambda_i). \quad (4.3)$$

Again, similar to Proposition 1, it can be shown that our distance $W_{2,\text{FR}}$ is the geodesic distance on $\mathcal{H}_{++}$ with the Riemannian-type metric (4.2).

Proposition 2

$$W_{2,\text{FR}}(\rho_0, \rho_1) = \inf_{\rho \in \mathcal{H}_{++}} \int_0^1 \sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}^{\text{FR}}} dt,$$

where the minimum is taken over all the continuously differentiable paths on the manifold $\mathcal{H}_{++}$ satisfying the boundary conditions $\rho(0) = \rho_0, \rho(1) = \rho_1$.

Proof The proof is the same as the proof of Proposition 1. □

Therefore, $W_{2,\text{FR}}(\cdot, \cdot)$ indeed defines a metric on $\mathcal{H}_{++}$ and it is intrinsic with respect to the Riemannian-type metric (4.2). We remark that even though the Riemannian-type metric (4.2) may not be well defined on the boundary of $\mathcal{H}_+$, $W_{2,\text{FR}}$ can be extended to $\mathcal{H}_+$ by the continuity with little effort [4]. If $\rho(\cdot)$ is a solution of (4.1), then it possesses the nice property

$$W_{2,\text{FR}}(\rho(s), \rho(t)) = (t - s) W_{2,\text{FR}}(\rho_0, \rho_1)$$

for all $0 \leq s < t \leq 1$. This is based on a simple argument using reparameterization. With this in mind, we now present a sufficient condition for the existence of geodesics. The equations (4.4) also characterize the geodesics.

Theorem 3 Suppose there exists $\lambda(\cdot) \in C^1([0, 1]; \mathcal{H})$ satisfying

$$\dot{\lambda} + \frac{1}{2} (\nabla_L \lambda)^* (\nabla_L \lambda) + \frac{1}{2\alpha} \lambda^2 = 0, \quad (4.4a)$$

and $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ satisfying

$$\dot{\rho} = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda) + \frac{1}{2\alpha} (\lambda \rho + \rho \lambda), \quad (4.4b)$$

and matching the marginals $\rho(0) = \rho_0, \rho(1) = \rho_1$. Then, the triple $(\rho, v, r) = (\rho, \nabla_L \lambda, \frac{1}{\alpha} \lambda)$ solves (4.1).

Proof Let $\lambda(\cdot) \in C^1([0, 1]; \mathcal{H})$ be a Lagrangian multiplier for the constraints (4.1b) and construct the Lagrangian

$$\begin{aligned}\mathcal{L}(\rho, v, r, \lambda) &= \int_0^1 \left\{ \frac{1}{2} \operatorname{tr}(\rho v^* v) + \frac{\alpha}{2} \operatorname{tr}(\rho r^* r) + \operatorname{tr}(\lambda(\dot{\rho} - \frac{1}{2} \nabla_L^* (v\rho + \rho v) - \frac{1}{2}(r\rho + \rho r))) \right\} dt \\ &= \int_0^1 \left\{ \frac{1}{2} \operatorname{tr}(\rho v^* v) - \frac{1}{2} \operatorname{tr}((\nabla_L \lambda)^* (v\rho + \rho v)) + \frac{\alpha}{2} \operatorname{tr}(\rho r^* r) \right. \\ &\quad \left. - \frac{1}{2} \operatorname{tr}(\lambda(r\rho + \rho r)) - \operatorname{tr}(\dot{\lambda}\rho) \right\} dt + \operatorname{tr}(\lambda(1)\rho_1) - \operatorname{tr}(\lambda(0)\rho_0).\end{aligned}$$

Pointwise minimizing the above over v, r yields

$$v_{\text{opt}}(t) = \nabla_L \lambda(t),$$

and

$$r_{\text{opt}}(t) = \frac{1}{\alpha} \dot{\lambda}(t).$$

The corresponding minimum is

$$\int_0^1 \left\{ -\frac{1}{2} \operatorname{tr}(\rho(\nabla_L \lambda)^* (\nabla_L \lambda)) - \frac{1}{2\alpha} \operatorname{tr}(\rho \dot{\lambda}^2) - \operatorname{tr}(\dot{\lambda}\rho) \right\} dt + \operatorname{tr}(\lambda(1)\rho_1) - \operatorname{tr}(\lambda(0)\rho_0).$$

When λ satisfies (4.4a), the first term of the above becomes 0 and therefore the triple $(\rho, v, r) = (\rho, \nabla_L \lambda, \frac{1}{\alpha} \dot{\lambda})$ is a minimizer of the Lagrangian $\mathcal{L}(\rho, v, r, \lambda)$. In view of (4.4b), it is also feasible. Therefore, it is a minimizer of (4.1). This completes the proof. $\square$

As in the balanced case [6], (4.1) has the following convex reformulation

$$\inf_{\rho \in \mathcal{H}_{++}, u \in \mathcal{C}^N, s \in \mathcal{C}} \int_0^1 \{ \operatorname{tr}(u\rho^{-1}u^*) + \alpha \operatorname{tr}(s\rho^{-1}s^*) \} dt, \quad (4.5a)$$

$$\dot{\rho} = \frac{1}{2} \nabla_L^* (u - \bar{u}) + \frac{1}{2} (s + s^*), \quad (4.5b)$$

$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1. \quad (4.5c)$$

Here, we simply used the change of variables $u = v\rho, s = r\rho$. The minimum is taken over all the density flows $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ and fluxes $u(\cdot) \in C([0, 1]; \mathcal{C}^N), s(\cdot) \in C([0, 1]; \mathcal{C})$. Note here we have relaxed the conditions on u, s as in general u, s cannot be written as $v\rho, s = r\rho$ for $v \in \mathcal{S}^N, r \in \mathcal{H}$. This relaxation, however, does not change the problem by the following lemma.

Lemma 4 For a given pair $\rho \in \mathcal{H}_{++}, \delta \in \mathcal{H}$, let $(u, s) = (\nabla_L \lambda \rho, \frac{1}{\alpha} \dot{\lambda} \rho)$ with λ being the unique solution to

$$\delta = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda) + \frac{1}{2\alpha} (\lambda \rho + \rho \lambda).$$

Then, it minimizes

$$\operatorname{tr}(u\rho^{-1}u^*) + \alpha \operatorname{tr}(s\rho^{-1}s^*)$$

over all $u \in \mathcal{C}^N, s \in \mathcal{C}$ such that

$$\delta = \frac{1}{2} \nabla_L^*(u - \bar{u}) + \frac{1}{2}(s + s^*).$$

Proof Consider the convex optimization problem

$$\min_{u \in \mathcal{C}^N, s \in \mathcal{C}} \frac{1}{2} \operatorname{tr}(u \rho^{-1} u^*) + \frac{\alpha}{2} \operatorname{tr}(s \rho^{-1} s^*),$$

$$\delta = \frac{1}{2} \nabla_L^*(u - \bar{u}) + \frac{1}{2}(s + s^*).$$

Obviously, it is feasible, so a minimizer exists. Besides, the cost function is strictly convex, thus the minimizer is unique. By Karush–Kuhn–Tucker conditions, there exists Lagrangian multiplier $\lambda \in \mathcal{H}$ such that the minimizer $(u_{\text{opt}}, s_{\text{opt}})$ minimizes the Lagrangian

$$\frac{1}{2} \operatorname{tr}(u \rho^{-1} u^*) + \frac{\alpha}{2} \operatorname{tr}(s \rho^{-1} s^*) + \langle \lambda, \delta - \frac{1}{2} \nabla_L^*(u - \bar{u}) - \frac{1}{2}(s + s^*) \rangle,$$

from which we conclude $(u_{\text{opt}}, s_{\text{opt}}) = (\nabla_L \lambda \rho, \frac{1}{\alpha} \lambda \rho)$. $\square$

Therefore, for any feasible solution (ρ, u, s) to (4.5), we can always construct $(\rho, \hat{u}, \hat{s}) = (\rho, \nabla_L \lambda \rho, \frac{1}{\alpha} \lambda \rho)$ such that the cost in (4.5a) stay unchanged.

Remark: In the scalar setting, the dynamic formulation of the interpolation between Wasserstein and Fish-Rao distance (2.6) can be rewritten in the static form (2.2) with a special cost function $c(\cdot, \cdot)$ [15]. This is not the cases in matrix setting. The dynamic formulation (4.1) does not have a static counterpart.

4.2 Interpolation between Wasserstein and Frobenius

As a straightforward generalization of (2.5), we define, for $\rho_0, \rho_1 \in \mathcal{H}_{++}$,

$$W_{2,F}(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{H}_{++}, v \in \mathcal{S}^N, s \in \mathcal{H}} \int_0^1 \{ \operatorname{tr}(\rho v^* v) + \alpha \operatorname{tr}(s^2) \} dt, \quad (4.6a)$$

$$\dot{\rho} = \frac{1}{2} \nabla_L^*(v \rho + \rho v) + s, \quad (4.6b)$$

$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1. \quad (4.6c)$$

Here, the minimum is taken over all the density flows $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ and velocity fields $v(\cdot) \in C([0, 1]; \mathcal{S}^N), s(\cdot) \in C([0, 1]; \mathcal{H})$. The second part of the cost corresponds to the Frobenius metric. More specifically, the Frobenius metric can be rewritten as

$$\|\rho_0 - \rho_1\|_F^2 = \inf_{\rho \in \mathcal{H}, s \in \mathcal{H}} \left\{ \int_0^1 \operatorname{tr}(s^2) dt \mid \dot{\rho} = s, \rho(0) = \rho_0, \rho(1) = \rho_1 \right\}.$$

Lemma 5 For a given pair $\rho \in \mathcal{H}_{++}, \delta \in \mathcal{H}$, there is a unique solution to

$$\delta = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda) + \frac{1}{\alpha} \lambda.$$

Moreover, $(v, s) = (\nabla_L \lambda, \frac{1}{\alpha} \lambda)$ is the unique minimizer of

$$\text{tr}(\rho v^* v) + \alpha \text{tr}(s^2)$$

over all the $(v \in \mathcal{S}^N, s \in \mathcal{H})$ satisfying

$$\delta = \frac{1}{2} \nabla_L^* (v \rho + \rho v) + s.$$

Proof The proof is omitted as it is almost identical to that of Lemma 2. □

Given any two tangent vectors δ_1, δ_2 at ρ on $\mathcal{H}_{++}$, the inner product

$$\langle \delta_1, \delta_2 \rangle_\rho^F = \frac{1}{2} \text{tr}(\rho (\nabla_L \lambda_1)^* \nabla_L \lambda_2 + \rho (\nabla_L \lambda_2)^* \nabla_L \lambda_1) + \frac{1}{\alpha} \text{tr}(\lambda_1 \lambda_2), \quad (4.7)$$

endows the manifold $\mathcal{H}_{++}$ a Riemannian-type structure. Here, $\lambda_i \in \mathcal{H}$, $(i = 1, 2)$ is the unique solution to

$$\delta_i = \frac{1}{2} \nabla_L^* (\nabla_L \lambda_i \rho + \rho \nabla_L \lambda_i) + \frac{1}{\alpha} \lambda_i. \quad (4.8)$$

We note that similar results as those described previously for $W_{2, \text{FR}}$, can be proven for the distance $W_{2, F}$. In particular, the latter gives the geodesic distance on $\mathcal{H}_{++}$ with respect to the Riemannian-type metric (4.7).

Proposition 3

$$W_{2, F}(\rho_0, \rho_1) = \inf_{\rho \in \mathcal{H}_{++}} \int_0^1 \sqrt{\langle \dot{\rho}(t), \dot{\rho}(t) \rangle_{\rho(t)}^F} dt,$$

where the minimum is taken over all the continuously differentiable paths on the manifold $\mathcal{H}_{++}$ satisfying the boundary conditions $\rho(0) = \rho_0, \rho(1) = \rho_1$.

Proof The proof is the same as the proof of Proposition 1. □

Employing a Lagrangian argument, we obtain the optimality condition and a characterization of the geodesics as follows.

Theorem 6 Suppose there exists $\lambda(\cdot) \in C^1([0, 1]; \mathcal{H})$ satisfying

$$\dot{\lambda} + \frac{1}{2} (\nabla_L \lambda)^* (\nabla_L \lambda) = 0 \quad (4.9a)$$

and $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ satisfying

$$\dot{\rho} = \frac{1}{2} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda) + \frac{1}{\alpha} \lambda \quad (4.9b)$$

and matching the marginals $\rho(0) = \rho_0, \rho(1) = \rho_1$. Then, the triple $(\rho, v, s) = (\rho, \nabla_L \lambda, \frac{1}{\alpha} \lambda)$ solves (4.6).

Further, through change of variable $u = v\rho$, we may reformulate (4.6) as the following convex optimization problem:

$$\inf_{\rho \in \mathcal{H}_{++}, u \in \mathcal{C}^N, s \in \mathcal{H}} \int_0^1 \{\text{tr}(u\rho^{-1}u^*) + \alpha \text{tr}(s^2)\} dt, \quad (4.10a)$$

$$\dot{\rho} = \frac{1}{2} \nabla_L^*(u - \bar{u}) + s, \quad (4.10b)$$

$$\rho(0) = \rho_0, \quad \rho(1) = \rho_1. \quad (4.10c)$$

The minimum is taken over all the density flows $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$ and fluxes $u(\cdot) \in C([0, 1]; \mathcal{C}^N)$, $s(\cdot) \in C([0, 1]; \mathcal{H})$. Again, the relaxation from $u = v\rho$ with $v \in S^N$ to general $u \in \mathcal{C}^N$ does not change the minimum by the following lemma.

Lemma 7 For a given pair $\rho \in \mathcal{H}_{++}, \delta \in \mathcal{H}$, let $(u, s) = (\nabla_L \lambda \rho, \frac{1}{\alpha} \lambda)$ with λ being the unique solution to

$$\delta = \frac{1}{2} \nabla_L^*(\nabla_L \lambda \rho + \rho \nabla_L \lambda) + \frac{1}{\alpha} \lambda.$$

Then, it minimizes

$$\text{tr}(u\rho^{-1}u^*) + \alpha \text{tr}(s^2)$$

over all $u \in \mathcal{C}^N, s \in \mathcal{H}$ such that

$$\delta = \frac{1}{2} \nabla_L^*(u - \bar{u}) + s.$$

Proof The proof is identical to that of Lemma 4. □

Remark: An alternative way to interpolate W_2 and the Frobenius metric is

$$\begin{aligned} \hat{W}_{2,F}(\rho_0, \rho_1)^2 &:= \inf_{\rho \in \mathcal{H}_{++}, \tilde{\rho}_1 \in \mathcal{H}_{++}, v \in S^N} \int_0^1 \text{tr}(\rho v^* v) dt + \alpha \|\rho_1 - \tilde{\rho}_1\|_F^2, \\ \dot{\rho} &= \frac{1}{2} \nabla_L^*(v\rho + \rho v), \\ \rho(0) &= \rho_0, \quad \rho(1) = \tilde{\rho}_1, \end{aligned}$$

where the minimum is taken over all $\tilde{\rho}_1 \in \mathcal{H}_{++}$, density flows $\rho(\cdot) \in C^1([0, 1]; \mathcal{H}_{++})$, and velocity fields $v(\cdot) \in C([0, 1]; S^N)$. This is a non-commutative analogue of (2.4). Note that in general $\hat{W}_{2,F}$ is not a metric.

5 Interpolation of the matrix-valued densities

In applications, it is often the case that one has to deal with the matrix-valued densities on dimensions which may represent space or frequency. Formally, we define a *spatial*

matrix-valued density to be a mapping $\rho : E \rightarrow \mathcal{H}_+$, where $E \subset \mathbb{R}^m$. For instance, in the context of multivariable time series analysis, it is natural to consider $m = 1$ and $E = [-\pi, \pi]$; see, e.g., [18]. For simplicity, we assume E to be a convex compact set.

Let

$$\mathcal{E} = \{\rho(\cdot) \mid \rho(x) \in \mathcal{H}_+ \text{ for } x \in E \text{ such that } \int_E \text{tr}(\rho(x)) dx < \infty\}, \quad (5.1)$$

and $\mathcal{E}_+$ its interior, namely,

$$\mathcal{E}_+ = \{\rho(\cdot) \mid \rho(x) \in \mathcal{H}_{++} \text{ for } x \in E \text{ such that } \int_E \text{tr}(\rho(x)) dx < \infty\}.$$

The tangent space at $\rho \in \mathcal{E}_+$ is

$$T_\rho = \{\delta(\cdot) \mid \delta(x) \in \mathcal{H} \text{ for } x \in E \text{ such that } \int_E |\text{tr}(\delta(x))| dx < \infty\}.$$

Note that the problem on the subset $\int_E \text{tr}(\rho(x)) dx = 1$, i.e., the balanced case, has been studied in [6]. By combining the standard continuity equation on the Euclidean space and the continuity equation for positive definite matrices (3.4), and taking into the source term, we obtain a continuity equation on $\mathcal{E}_+$ for the flow $\rho(t, x)$ as

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w\rho + \rho w) - \frac{1}{2} \nabla_L^* (v\rho + \rho v) - \frac{1}{2} (r\rho + \rho r) = 0, \quad (5.2)$$

or simply

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w\rho + \rho w) - \frac{1}{2} \nabla_L^* (v\rho + \rho v) - s = 0. \quad (5.3)$$

Here, $\nabla_x \cdot$, the standard divergence operator on $\mathbb{R}^m$ applies independently to each entry of $w\rho + \rho w$, $w(t, x) \in \mathcal{H}^m$ is the velocity field along the spacial dimension, and $v(t, x) \in S^N$ is the quantum velocity as before. The source term is $\frac{1}{2}(r\rho + \rho r)$ in (5.2) and s in (5.3). We next present, based on the continuity equations, both the interpolating distance between Wasserstein and Fisher-Rao, and the interpolating distance between Wasserstein and Frobenius metric.

5.1 Interpolation between Wasserstein and Fisher-Rao

A dynamic formulation of matrix-valued OMT between two given marginals $\rho_0, \rho_1 \in \mathcal{E}_+$ is

$$W_{2,\text{FR}}(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{E}_+, w \in \mathcal{H}^m, v \in S^N, r \in \mathcal{H}} \int_0^1 \int_E \{ \text{tr}(\rho w^* w) + \gamma \text{tr}(\rho v^* v) + \alpha \text{tr}(\rho r^2) \} dx dt, \quad (5.4a)$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w\rho + \rho w) - \frac{1}{2} \nabla_L^* (v\rho + \rho v) - \frac{1}{2} (r\rho + \rho r) = 0, \quad (5.4b)$$

$$\rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1. \quad (5.4c)$$

Note that the minimum is taken over all the smooth density flows ρ and velocity fields w, v , and r with the zero flux boundary conditions $(w\rho + \rho w) \cdot \vec{n} = 0$ on ∂E with $\vec{n}$

denoting the normal vector of ∂E . The coefficient $\gamma > 0$ is arbitrary and weighs in the relative significance of the two velocity fields w, v . We then define the interpolating distance $W_{2,FR}(\rho_0, \rho_1)$ between ρ_0 and ρ_1 via (5.4 a).

The proof of the existence of minimum of (5.4) is not the focus of this paper and will be studied somewhere else. Here, we provide a sufficient condition for optimality.

Theorem 8 Suppose there exists smooth $\lambda(\cdot, \cdot) \in \mathcal{H}$ satisfying

$$\frac{\partial \lambda}{\partial t} + \frac{1}{2}(\nabla_x \lambda)^*(\nabla_x \lambda) + \frac{1}{2\gamma}(\nabla_L \lambda)^*(\nabla_L \lambda) + \frac{1}{2\alpha} \lambda^2 = 0 \quad (5.5a)$$

and smooth $\rho(\cdot) \in \mathcal{E}_+$ satisfying

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (\nabla_x \lambda \rho + \rho \nabla_x \lambda) - \frac{1}{2\gamma} \nabla_L^* (\nabla_L \lambda \rho + \rho \nabla_L \lambda) - \frac{1}{2\alpha} (\lambda \rho + \rho \lambda) = 0 \quad (5.5b)$$

and matching the two marginals $\rho(0, \cdot) = \rho_0, \rho(1, \cdot) = \rho_1$. If in addition, $(\nabla_x \lambda \rho + \rho \nabla_x \lambda) \cdot \vec{n} = 0$ on ∂E , then $(\rho, w, v, r) = (\rho, \nabla_x \lambda, \frac{1}{\gamma} \nabla_L \lambda, \frac{1}{\alpha} \lambda)$ solves (5.4).

Proof Let $\lambda(\cdot, \cdot) \in \mathcal{H}$ be a smooth function and define the Lagrangian

$$\begin{aligned} \mathcal{L}(\rho, v, w, \lambda) = & \int_0^1 \int_E \left\{ \frac{1}{2} \text{tr}(\rho w^* w) + \frac{\gamma}{2} \text{tr}(\rho v^* v) + \frac{\alpha}{2} \text{tr}(\rho r^2) \right. \\ & \left. + \text{tr}(\lambda (\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w \rho + \rho w) - \frac{1}{2} \nabla_L^* (v \rho + \rho v) - \frac{1}{2} (r \rho + \rho r))) \right\} dx dt. \end{aligned}$$

Integration by parts, in view of the boundary conditions on w , yields

$$\begin{aligned} & \int_0^1 \int_E \left\{ \frac{1}{2} \text{tr}(\rho w^* w) + \frac{\gamma}{2} \text{tr}(\rho v^* v) + \frac{\alpha}{2} \text{tr}(\rho r^2) \right. \\ & \left. - \text{tr}(\lambda \frac{\partial \rho}{\partial t}) - \frac{1}{2} \langle \nabla_x \lambda, w \rho + \rho w \rangle - \frac{1}{2} \langle \nabla_L \lambda, v \rho + \rho v \rangle - \frac{1}{2} \text{tr}(\lambda (r \rho + \rho r)) \right\} dx dt. \end{aligned}$$

Here, we have discarded the terms on ρ_0, ρ_1 . Minimizing the above pointwise over w, v gives expressions for the optimal values as

$$w_{\text{opt}}(t, x) = \nabla_x \lambda(t, x),$$

$$v_{\text{opt}}(t, x) = \frac{1}{\gamma} \nabla_L \lambda(t, x),$$

and

$$r_{\text{opt}}(t, x) = \frac{1}{\alpha} \lambda(t, x).$$

Substituting these back to the Lagrangian, we obtain

$$\int_0^1 \int_E \left\{ -\frac{1}{2} \text{tr}(\rho (\nabla_x \lambda)^*(\nabla_x \lambda)) - \frac{1}{2\gamma} \text{tr}(\rho (\nabla_L \lambda)^*(\nabla_L \lambda)) - \frac{1}{2\alpha} \text{tr}(\rho \lambda^2) - \text{tr}(\rho \frac{\partial \lambda}{\partial t}) \right\} dx dt.$$

Hence, $(\rho, w, v, r) = (\rho, \nabla_x \lambda, \frac{1}{\gamma} \nabla_L \lambda, \frac{1}{\alpha} \lambda)$ minimizes the Lagrangian $\mathcal{L}$. By (5.5b), it is also feasible. Therefore, it solves (5.4). $\square$

Formally, the Wasserstein distance $W_{2,\text{FR}}(\rho, \rho + \delta)$ induces a Riemannian-type structure on the tangent space of T_ρ . Given any two tangent vector δ_1, δ_2 at ρ , we can associate them with λ_1, λ_2 by solving

$$\delta_j = -\frac{1}{2} \nabla_x \cdot (\nabla_x \lambda_j \rho + \rho \nabla_x \lambda_j) + \frac{1}{2\gamma} \nabla_L^* (\nabla_L \lambda_j \rho + \rho \nabla_L \lambda_j) + \frac{1}{2\alpha} (\lambda_j \rho + \rho \lambda_j), \quad j = 1, 2, \quad (5.6)$$

and define the Riemannian-type metric as

$$\begin{aligned} \langle \delta_1, \delta_2 \rangle_\rho^{\text{FR}} = & \int_E \left\{ \frac{1}{2} \text{tr}(\rho (\nabla_x \lambda_1)^* \nabla_x \lambda_2 + \rho (\nabla_x \lambda_2)^* \nabla_x \lambda_1) \right. \\ & \left. + \frac{1}{2\gamma} \text{tr}(\rho (\nabla_L \lambda_1)^* \nabla_L \lambda_2 + \rho (\nabla_L \lambda_2)^* \nabla_L \lambda_1) + \frac{1}{2\alpha} \text{tr}(\rho \lambda_1 \lambda_2 + \rho \lambda_2 \lambda_1) \right\} dx. \end{aligned}$$

Again, $W_{2,\text{FR}}(\cdot, \cdot)$ is a metric on $\mathcal{E}_+$ and it is intrinsic. That is,

$$W_{2,\text{FR}} = \inf_{\rho} \int_0^1 \sqrt{\left\langle \frac{\partial \rho}{\partial t}, \frac{\partial \rho}{\partial t} \right\rangle_{\rho(t)}^{\text{FR}}} dt,$$

where the integral is minimized over all the piecewise smooth curves in $\mathcal{E}_+$ connecting ρ_0 and ρ_1 . The proof follows the same spirit as that of Proposition 1, but it requires much more delicate analysis as we now need to work on partial differential equations instead of differential equations. Rigorous proof will be provided somewhere else.

Finally, let $q = w\rho, u = v\rho, s = r\rho$, then (5.4) can again be cast as a convex optimization problem

$$\inf_{\rho \in \mathcal{E}_+, q \in \mathcal{C}^m, u \in \mathcal{C}^N, s \in \mathcal{C}} \int_0^1 \int_E \{ \text{tr}(q\rho^{-1}q^*) + \gamma \text{tr}(u\rho^{-1}u^*) + \alpha \text{tr}(s\rho^{-1}s^*) \} dx dt, \quad (5.7a)$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (q + \bar{q}) - \frac{1}{2} \nabla_L^* (u - \bar{u}) - \frac{1}{2} (s + s^*) = 0, \quad (5.7b)$$

$$\rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1. \quad (5.7c)$$

Note that the minimum is taken over all smooth functions with zero flux boundary conditions on q .

5.2 Interpolation between Wasserstein and Frobenius

Given $\rho_0, \rho_1 \in \mathcal{E}_+$, we define the interpolating distance between Wasserstein and Frobenius through

$$W_{2,F}(\rho_0, \rho_1)^2 := \inf_{\rho \in \mathcal{E}_+, w \in \mathcal{H}^m, v \in \mathcal{S}^N, s \in \mathcal{H}} \int_0^1 \int_E \{ \text{tr}(\rho w^* w) + \gamma \text{tr}(\rho v^* v) + \alpha \text{tr}(s^2) \} dx dt, \quad (5.8a)$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w\rho + \rho w) - \frac{1}{2} \nabla_L^* (v\rho + \rho v) - s = 0, \quad (5.8b)$$

$$\rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1. \quad (5.8c)$$

The minimum is taken over all the piecewise smooth density flows ρ and velocity fields w, v , and r with the zero flux boundary conditions $(w\rho + \rho w) \cdot \vec{n} = 0$ on ∂E . Following similar argument as in Theorem 8, we arrive at an optimality condition.

Theorem 9 Suppose there exists smooth $\lambda(\cdot, \cdot) \in \mathcal{H}$ satisfying

$$\frac{\partial \lambda}{\partial t} + \frac{1}{2} (\nabla_x \lambda)^* (\nabla_x \lambda) + \frac{1}{2\gamma} (\nabla_L \lambda)^* (\nabla_L \lambda) = 0, \quad (5.9a)$$

and smooth $\rho(\cdot) \in \mathcal{E}_+$ satisfying

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (\nabla_x \lambda \rho + \rho \nabla_x \lambda) - \frac{1}{2\gamma} \nabla_L^* (\rho \nabla_L \lambda + \nabla_L \lambda \rho) - \frac{1}{\alpha} \lambda = 0, \quad (5.9b)$$

and matching the two marginals $\rho(0, \cdot) = \rho_0, \rho(1, \cdot) = \rho_1$. If in addition, $(\nabla_x \lambda \rho + \rho \nabla_x \lambda) \cdot \vec{n} = 0$ on ∂E , then $(\rho, w, v, s) = (\rho, \nabla_x \lambda, \frac{1}{\gamma} \nabla_L \lambda, \frac{1}{\alpha} \lambda)$ solves (5.8).

The Wasserstein distance $W_{2,F}(\rho, \rho + \delta)$ induces a Riemannian-type structure on the tangent space of T_ρ . Given any two tangent vector δ_1, δ_2 at ρ , we can associate them with λ_1, λ_2 by solving

$$\delta_j = -\frac{1}{2} \nabla_x \cdot (\nabla_x \lambda_j \rho + \rho \nabla_x \lambda_j) + \frac{1}{2\gamma} \nabla_L^* (\nabla_L \lambda_j \rho + \rho \nabla_L \lambda_j) + \frac{1}{\alpha} \lambda_j, \quad j = 1, 2, \quad (5.10)$$

and define the Riemannian-type metric as

$$\begin{aligned} \langle \delta_1, \delta_2 \rangle_\rho^F = & \int_E \left\{ \frac{1}{2} \text{tr}(\rho (\nabla_x \lambda_1)^* \nabla_x \lambda_2 + \rho (\nabla_x \lambda_2)^* \nabla_x \lambda_1) \right. \\ & \left. + \frac{1}{2\gamma} \text{tr}(\rho (\nabla_L \lambda_1)^* \nabla_L \lambda_2 + \rho (\nabla_L \lambda_2)^* \nabla_L \lambda_1) + \frac{1}{\alpha} \text{tr}(\lambda_1 \lambda_2) \right\} dx. \end{aligned}$$

The metric $W_{2,F}$ is intrinsic, namely,

$$W_{2,F} = \inf_\rho \int_0^1 \sqrt{\left\langle \frac{\partial \rho}{\partial t}, \frac{\partial \rho}{\partial t} \right\rangle_{\rho(t)}^F} dt,$$

where the integral is minimized over all the piecewise smooth curves in $\mathcal{E}_+$ connecting ρ_0 and ρ_1 .

Substituting $q = w\rho$ and $u = v\rho$, we get a convex formulation

$$\inf_{\rho \in \mathcal{E}_+, q \in \mathcal{C}^m, u \in \mathcal{C}^N, s \in \mathcal{H}} \int_0^1 \int_E \{ \text{tr}(q\rho^{-1}q^*) + \gamma \text{tr}(u\rho^{-1}u^*) + \alpha \text{tr}(s^2) \} dx dt, \quad (5.11a)$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (q + \bar{q}) - \frac{1}{2} \nabla_L^*(u - \bar{u}) - s = 0, \quad (5.11b)$$

$$\rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1. \quad (5.11c)$$

The minimum is taken over smooth functions with zero flux boundary conditions on q .

6 Examples

One typical example of the matrix-valued density is the power spectrum density of a multivariate stochastic signal. The power spectrum of a stationary discrete time stochastic process is a map from $E = [-\pi, \pi]$ to $\mathcal{H}$. The dimension of $\mathcal{H}$ is equal to the number of channels of the signal. The power spectrum describe the power distribution of the signal among different frequencies as well as the correlation between channels.

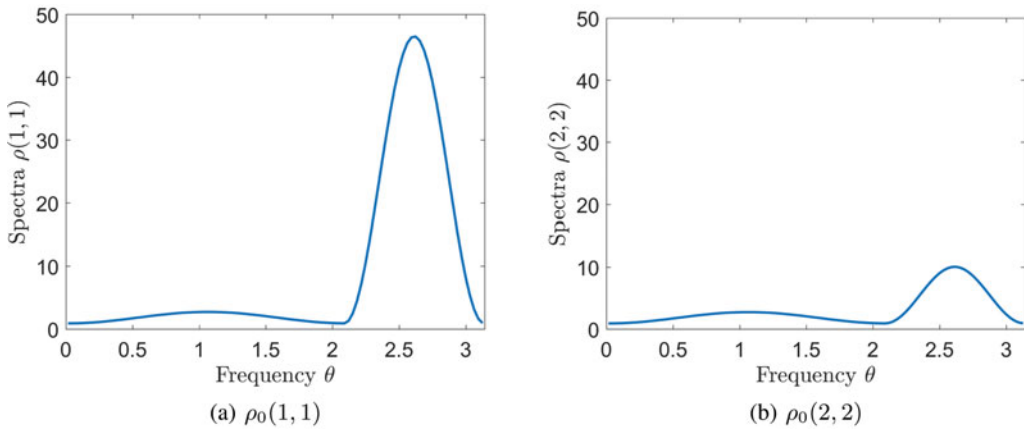
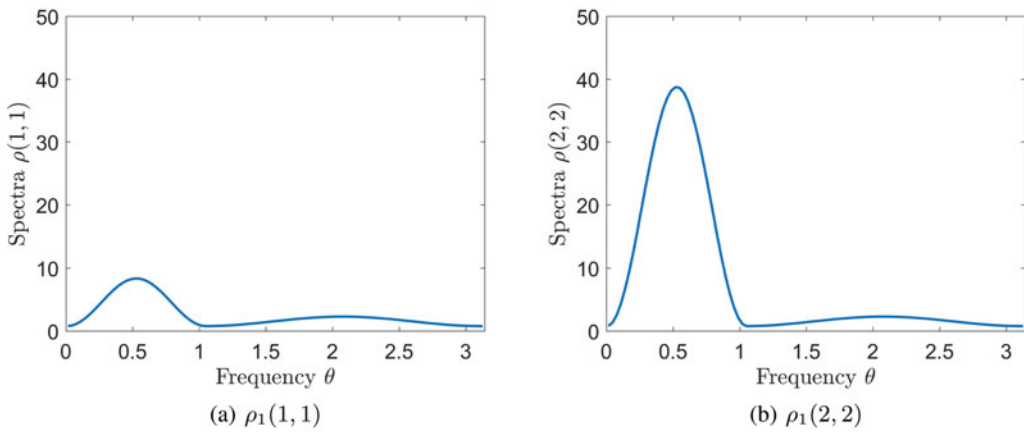
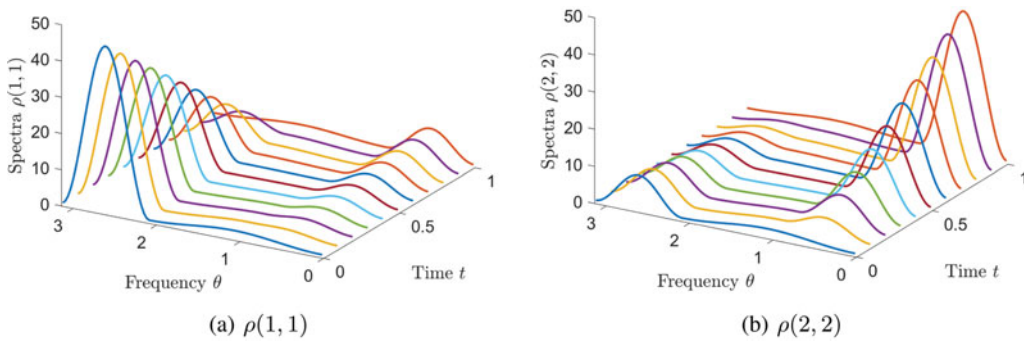
When enough data samples are given, one can estimate the power spectrum using Fourier transform or other similar tools. This is, however, built on the assumption that the signal is stationary. When the signal is time-varying, its power spectrum will change as a function of time. This is often the case. To capture the varying of power spectrum, we can use Fourier transform on small time windows to estimate the power spectrum at different time points. To connect them through a continuous function of power spectrum, we need a good method to interpolate these spectra. Next, we show that the interpolation using our framework makes perfect sense.

Consider two power spectra of two-dimensional signals as shown in Figures 1 and 2. If we use simply linear interpolation, then the result is shown in Figure 3. It has fade-in fade-out effect and cannot capture the modalities of the signal. On the other hand, if we use geodesic interpolation with respect $W_{2,FR}$ or $W_{2,F}$, we get the spectra flow in Figures 4 and 5, respectively. We can see that the dominant frequencies change slowly from one to another. In such a way, we are able to capture the moving of the energy peaks in the signals.

7 Conclusions

Our line of research into unbalanced versions of OMT is motivated by the fact that general distributions (histograms, power spectra, spatio-temporal energy densities, and images) may not necessarily be normalized to have the same integral. Thus, it is imperative to devise appropriate metrics and theory to handle these situations. Our overall aim is to provide constructions for ‘interpolating’ data in the form of distributions. In the present work, we have formulated a natural technique that interpolates the quantum mechanical version of OMT developed in [6] with an analogue of Fisher-Rao information.

In the future work, we plan to establish the existence proofs for our general optimal transport problems. To achieve this goal, we will focus on the convex reformulations with weaker assumptions, and follow a similar argument as in [5]. We will also explore the

FIGURE 1. ρ_0 . (a) $\rho_0(1,1)$ and (b) $\rho_0(2,2)$.FIGURE 2. ρ_1 . (a) $\rho_1(1,1)$ and (b) $\rho_1(2,2)$.FIGURE 3. Linear interpolation. (a) $\rho(1,1)$ and (b) $\rho(2,2)$.

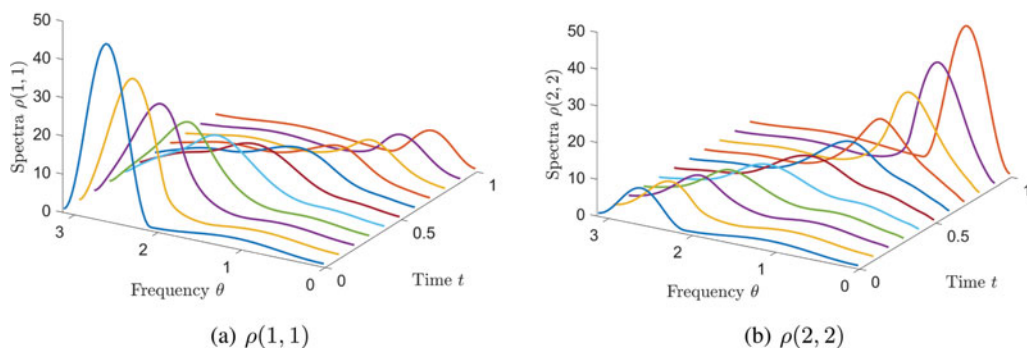


FIGURE 4. Interpolation based on $W_{2,FR}$. (a) $\rho(1, 1)$ and (b) $\rho(2, 2)$.

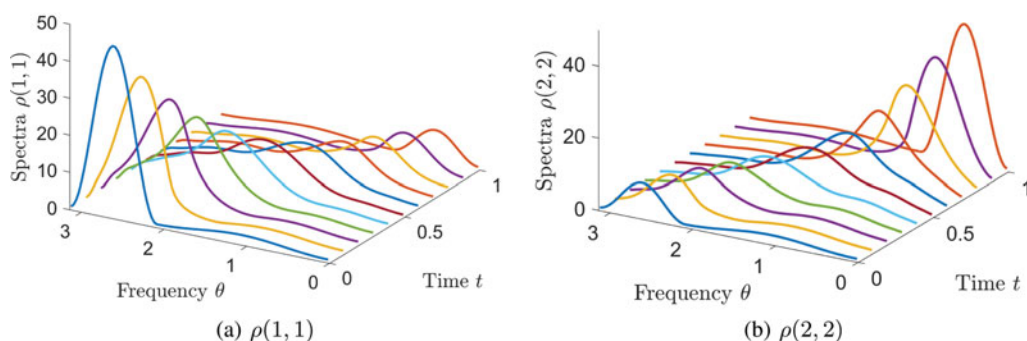


FIGURE 5. Interpolation based on $W_{2,F}$. (a) $\rho(1, 1)$ and (b) $\rho(2, 2)$.

Riemannian-type structure associated to the unbalanced Wasserstein distance and other variants of the continuity equation. From a more applied side, we plan to apply the methodology described in the present work to the various types of multi-modal, multi-sensor, and multi-spectral data. It seems ideal for multiple target estimation as was done in [18].

References

- [1] BENAMOU, J.-D. (2010) Numerical resolution of an unbalanced mass transport problem. *ESAIM: Math. Model. Numer. Anal.* **37**(2), 851–868.
- [2] BENAMOU, J.-D. & BRENIER, Y. (2000) A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84**, 375–393.
- [3] BOYD, S. & VANDENBERGHE, L. (2004) *Convex Optimization*, Cambridge University Press, Cambridge, England.
- [4] CARLEN, E. & MAAS, J. (2017) Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.* **273**, 1810–1869. <https://arxiv.org/abs/1609.01254>.
- [5] CHEN, Y., GANGBO, W., GEORGIU, T. T. & TANNENBAUM, A. (2017) On the matrix Monge–Kantorovich problem. Preprint arXiv:1701.02826.

- [6] CHEN, Y., GEORGIU, T. T. & TANNENBAUM, A. (2017) Matrix optimal mass transport: A quantum mechanical approach. *IEEE Trans. Autom. Control* (99), DOI: 10.1109/TAC.2017.2767707.
- [7] CHEN, Y., GEORGIU, T. T. & PAVON, M. (2016) On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *J. Optim. Theory Appl.* **169**(2), 671–691.
- [8] CHEN, Y., HABER, E., YAMAMOTO, K., GEORGIU, T. T. & TANNENBAUM, A. (2018) An efficient algorithm for matrix-valued and vector-valued optimal mass transport. *J. Sci. Comput.*, <https://doi.org/10.1007/s10915-018-0696-8>.
- [9] CHIZAT, L., SCHMITZER, B., PEYRÉ, G. & VIALARD, F.-X. (2018) An interpolating distance between optimal transport and Fisher-Rao. *Found. Comp. Math.* **18**, 1–44. <https://arxiv.org/pdf/1506.06430.pdf>.
- [10] GUSTAFSON, S. & SIGAL, I. M. (2011) *Mathematical Concepts of Quantum Mechanics*, Springer, New York.
- [11] JORDAN, R., KINDERLEHRER, D. & OTTO, F. (1998) The variational formulation of the Fokker–Planck equation *SIAM J. Math. Anal.* **29**, 1–17.
- [12] KANTOROVICH, L. V. (1948) On a problem of Monge. *Uspekhi Mat. Nauk.* **3**, 225–226.
- [13] KONDRATYEV, S., MONSANGEON, L. & VOROTNIKOV, D. (2016) A new optimal transport distance on the space of finite Radon measures. *Adv. Differ. Equ.* **21**, 1117–1164.
- [14] 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.
- [15] LIERO, M., MIELKE, A. & SAVARÉ, G. (2018) Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Invent. Math.* **211**, 969–1117. <https://arxiv.org/abs/1508.07941>.
- [16] MITTENZWEIG, M. & MIELKE, A. (2017) An entropic gradient structure for Lindblad equations and couplings of quantum systems to macroscopic models. *J. Stat. Phys.* **167**, 205–233. <https://arxiv.org/abs/1609.05765>.
- [17] MUELLER, M., KARASEV, P., KOLESOV, I. & TANNENBAUM, A. (2013) Optical flow estimation for flame detection in videos. *IEEE Trans. Image Process.* **22**(2), 2786–2797.
- [18] NING, L., GEORGIU, T. & TANNENBAUM, A. (2015) On matrix-valued Monge–Kantorovich optimal mass transport. *IEEE Trans. Autom. Control* **60**(2), 373–382.
- [19] RACHEV, S. & RÜSCHENDORF, L. (1998) *Mass Transportation Problems*, Volumes I and II, Probability and Its Applications, Springer, New York.
- [20] TANNENBAUM, E., GEORGIU, T. & TANNENBAUM, A. (2010) Signals and control aspects of optimal mass transport and the Boltzmann entropy. In: *Proceedings of the 49th IEEE Conf. Decision Control (CDC)*.
- [21] UHLMANN, A. (1992) The metric of Bures and the geometric phase. In: *Quantum Groups and Related Topics*, edited by R. Gielerak, Kluwer Academic Publishers, Dordrecht, pp. 267–274.
- [22] VILLANI, C. (2003) *Topics in Optimal Transportation*, Graduate Studies in Mathematics, vol. 58, AMS, Providence, RI.