Matrix Optimal Mass Transport: A Quantum Mechanical Approach

Yongxin Chen, Tryphon T. Georgiou, and Allen Tannenbaum

***Abstract*—In this paper, we describe a possible generalization of the Wasserstein-2 metric, originally defined on the space of scalar probability densities, to the space of Hermitian matrices with trace one and to the space of matrix-valued probability densities. Our approach follows a control-theoretic optimization formulation of the Wasserstein-2 metric, having its roots in fluid dynamics, and utilizes certain results from the quantum mechanics of open systems, in particular the Lindblad equation. It allows determining the gradient flow for the quantum entropy relative to this matricial Wasserstein metric.**

***Index Terms*—Matrix optimal transport, quantum mechanics, entropic flows, non-commutative Wasserstein.**

I. Introduction

Optimal mass transport (OMT) is a rich area of research with applications to numerous disciplines including automatic control, transportation, econometrics, fluid dynamics, statistical physics, shape optimization, expert systems, and meteorology; see [1] and [2] for extensive lists of references. The original problem was first formulated by Monge in 1781 and concerned finding the optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one site to another. Much later, the problem was extensively analyzed by Kantorovich [3] with a focus on economic resource allocation and so is now known as the Monge–Kantorovich or OMT problem.

In this paper, we develop a noncommutative counterpart of optimal transport where density matrices ρ (i.e., Hermitian matrices that are positive definite and have unit trace) replace probability distributions, and where "transport" corresponds to a flow on the space of such matrices that minimizes a corresponding action integral. In some recent work [4], to which we refer for control and signal processing applications, a certain approach was formulated that had its basis on Kantorovich's idea of regularization on a joint distribution in a suitable product space.

Manuscript received January 18, 2017; revised May 7, 2017 and July 28, 2017; accepted October 23, 2017. Date of publication October 30, 2017; date of current version July 26, 2018. This work was supported by the Air Force Office of Scientific Research under Grant FA9550-15-1-0045 and Grant FA9550-17-1-0435, by the ARO under Grant W911NF-17-1-0429, by the National Center for Research Resources under Grant P41-RR-013218, by the National Institute of Biomedical Imaging and Bioengineering under Grant P41-EB-015902, by the National Science Foundation, and by the National Institutes of Health under Grant P30-CA-008748. Recommended by Associate Editor K. Kashima. *(Corresponding author: Tryphon T. Georgiou.)*

Y. Chen is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011, USA (e-mail: yongchen@iastate.edu).

T. T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697, USA (e-mail: tryphon@uci.edu).

A. Tannenbaum is with the Department of Computer Science and the Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: allen.tannenbaum@stonybrook.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2017.2767707

In contrast, in the present work, we employ generalizations of the seminal approach of Benamou and Brenier [5]. This latter work gives rise to a control-theoretic optimization framework to transport theory. To this end, in the present paper, we combine the control approach of [5] with ideas from quantum mechanics [6] to derive version of noncommutative transport theory, which may be regarded as a control version of matrix-valued OMT.

In general, OMT induces a geometry on spaces of distributions, which is relevant to a number of real-world applications [1], [2]. Most notably, it induces the so-called Wasserstein-1 and Wasserstein-2 metrics, which correspond to the transport cost taken as the distance or the square of the distance, respectively. One of our goals in seeking theory for matrix-valued OMT (herein, one that is analogous to the Wasserstein-2 geometry) is in modeling the dynamical evolution of matrix-valued measures, e.g., power distributions of vector-valued time series [4]. For instance, the entries of corresponding time series may represent measurements of different modalities across an array of sensors (e.g., frequency/color, polarization, spatial characteristics, and other target attributes). Alternatively, they may simply represent entries of a vectorial output process in the context of system identification. Thus, we expect the present work to also have applications in multivariable time series analysis, system identification, as well as in quantum control and quantum information.

At about the same time as the present work was originally reported in [7], closely related approaches were formulated independently and simultaneously in [8] and [9]. While all three [7]–[9] begin with the Lindblad equation, [7] differs in several important aspects. First, in [7], presents several alternative options for the multiplication operator in the quantum mechanical continuity equation; see (15), (16b), and (17b). In contrast, the authors of [8] and [9] only consider (17b), which leads to a linear heat equation for the gradient flow of the quantum entropy. While we discovered this connection as well, our emphasis has been on the anticommutator version (16b), which has the distinct advantage of leading to *computable convex optimization for the corresponding Wasserstein metric*. Since we are interested in engineering applications, this is obviously crucial. Furthermore, we also study the case of spatially dependent densities and develop a matrix OMT theory for this as well, not dealt with in either [8] or [9]. Spatially dependent densities are important in applications—areas we have in mind include diffusion tensor imaging (DTI) in which one encounters spatially varying tensor fields and spectral analysis of multivariable time series, in which one may need to interpolate, average/fuse, or regularize matrix-valued power spectral distributions [4].

The framework that is introduced in the present paper has led to further subsequent developments [10], [11]. In particular, Chen *et al.* [11] introduce an alternative matrix-valued Wasserstein-1 distance, whereas Chen *et al.* [10] explore the possibility of relating matrix-valued distributions of unequal mass/trace (unbalanced). We point out that there are fundamental differences and respective advantages between the Wasserstein-1 theory (presented in [11]) and the Wasserstein-2 theory presented herein. In particular, Wasserstein-2 induces a Riemannian structure and a unique geodesic between matrices, which is not the case in Wasserstein-1.

II. Background on Optimal Transport

In this section, we highlight concepts and constructs from the theory of OMT for scalar-valued distributions (see [2] for details). We focus

on the fluid dynamical formulation of Benamou and Brenier and the Riemannian structure on the space of densities that originates in the work of Otto and his coworkers. This background section is sketchy but will allow us to draw analogies with the matrix-valued counterpart that follows.

A. Control-Theoretic Approach to OMT

Here, the ρ s represent positive distributions (density functions[1]) on a space such as $\mathbb{R}^m$ (assumed throughout this section). It was shown in [5] that the Monge–Kantorovich problem [2] with a quadratic cost, i.e., the problem to transfer mass which is initially distributed according to ρ_0 to a final distribution[2] ρ_1, optimally via a transfer map $x \mapsto T(x)$ that minimizes the cost, may be given a computational fluid formulation.

Indeed, the mass transfer cost, which is referred to as the *Wasserstein distance* W_2 between the two densities ρ_0 and ρ_1,

$$W_2(\rho_0,\rho_1)^2 := \inf_T \left\{ \int \|x - T(x)\|^2 \rho_0(x) dx \mid \rho_0(x) = \rho_1(T(x)) \det(\nabla T(x)) \right\}$$

can also be expressed as the infimum of the "action integral"

$$\inf \int \int_0^1 \rho(t,x) \|v(t,x)\|^2 \, dt \, dx \tag{1}$$

over all flows of time-varying densities $\rho(t,x)$ and velocity fields $v(t,x) \in \mathbb{R}^m$ (weakly) satisfying the *continuity equation*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0 \tag{2}$$

and the boundary conditions

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

This idea was transformative[3] in that it brought forth an elegant structure as a control problem to minimize kinetic energy by suitable choice of the control (velocity) field v. The optimal solution may be characterized by the following condition.

Theorem 1: The solution of the OMT problem (1) is

$$v(t,x) = \nabla \phi(t,x) \tag{3a}$$

where ϕ and the corresponding flow ρ satisfy

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} \|\nabla \phi\|^2 = 0 \tag{3b}$$

and

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \nabla \phi) = 0. \tag{3c}$$

It is easy to recognize (3b) as a Hamilton–Jacobi equation and (3c) as a Fokker–Planck equation. Also, it turns out that the functional (1) can be conveniently expressed as

$$\inf \int \int_0^1 \rho(t,x)^{-1} \|u(t,x)\|^2 \, dt \, dx \tag{4}$$

with $u = \rho v$ being a momentum field, which is convex with respect to the pair (ρ, u). Under fairly general conditions, the infimum is attained and the minimizing velocity field v is unique. Moreover, the minimizing velocity field is simply $x + t(\nabla \varphi(x) - x)$, where φ is a convex function and $T(x) = \nabla \varphi(x)$ is precisely the solution to the Monge–Kantorovich problem [2]. Thus, the analysis in [5], with the introduction of the action integral, provides a *physically motivated dynamical reinterpretation of the Monge–Kantorovich problem*.

B. Riemannian Manifold Structure on Scalar Probability Densities

Intuition into the problems we consider may be provided by considering certain physical insights. Indeed, there is an amazing connection between entropy functionals, the heat equation, and the geometry induced by the Wasserstein distance that has emerged in recent years [14]. We now briefly touch upon these as it will allow to draw analogies in the matricial setting that follows.

Consider the manifold of densities on $\mathbb{R}^m$ integrating[4] to 1:

$$\mathcal{D} := \left\{ \rho \geq 0 \mid \int \rho = 1 \right\}.$$

The tangent space at a given point ρ may be identified with functions δ integrating to 0:

$$T_\rho \cong \left\{ \delta \mid \int \delta = 0 \right\}.$$

The manifold $\mathcal{D}$ admits a Riemannian-type structure that induces the Wasserstein distance. The key idea essentially originated in [15] and was developed into a powerful geometric approach to OMT by Otto in [14]; see also [2] and [16].

More specifically, under suitable assumptions on differentiability for $\rho \in \mathcal{D}$ and $\delta \in T_\rho$, one solves the Poisson equation

$$\delta = -\nabla \cdot (\rho \nabla g). \tag{5}$$

This allows identifying elements δ in the tangent space with functions g, up to additive constant; thus, given δ, we denote the solution of (5) by g_δ and the corresponding vector field by $v_\delta := \nabla g_\delta$. Then, given $\delta_1, \delta_2 \in T_\rho$, we can define the inner product

$$\langle \delta_1, \delta_2 \rangle_\rho := \int \rho \langle v_{\delta_1}, v_{\delta_2} \rangle \tag{6}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^m$. An integration by parts argument shows that this inner product will exactly induce the Wasserstein distance $W_2(\rho_0,\rho_1)$ given by (1). Thus, given two "points" $\rho_0, \rho_1 \in \mathcal{D}$, the minimizer of the Benamou–Brenier formulation, which coincides with the *displacement interpolating* curve [17] between the two densities, $\rho(t,\cdot)$, is precisely a Wasserstein geodesic. Using integration by parts, we obtain

$$\begin{aligned} \|\delta\|_\rho^2 = \ \langle \delta, \delta \rangle_\rho &= \int \rho \nabla g_\delta \cdot \nabla g_\delta \\ &= -\int g_\delta \nabla \cdot (\rho \nabla g_\delta) = \int \delta g_\delta. \end{aligned} \tag{7}$$

The distance between ρ_0 and ρ_1 may be rewritten as

$$\begin{aligned} W_2(\rho_0,\rho_1) &= \min_\rho \int_0^1 \|\dot\rho(t)\|_{\rho(t)} dt \\ &= \min_\rho \int_0^1 \sqrt{\langle \dot\rho(t), \dot\rho(t) \rangle_{\rho(t)}} dt \end{aligned}$$

where the minimum is taken over all the piecewise smooth curves connecting ρ_0 and ρ_1 on the manifold $\mathcal{D}$.

[1] More generally, one may consider positive measures, but this is avoided for simplicity and ease of correspondence with the matrix case.

[2] The two marginals ρ_0 and ρ_1 are assumed to have finite second-order moments.

[3] The approach has led to several new directions in stochastic control. See [12] and [13] and the references therein.

[4] For brevity of notation, whenever there is no risk of confusion, the volume element dx is omitted.

C. Gradient Flow of the Entropy

We close by sketching the fact that the gradient flow with respect to the Wasserstein geometry of the entropy functional

$$S(\rho) = -\int \rho \log \rho$$

with $\rho \in \mathcal{D}$, is given by the heat equation [14], [16]. Indeed, evaluating S along a one-parameter family in $\mathcal{D}$, $\rho(t,\cdot)$, and taking the derivative with respect to t, in view of $\int \rho = 1$, we obtain

$$\frac{dS}{dt} = -\int \left(\frac{\partial \rho}{\partial t}\log\rho + \frac{\partial \rho}{\partial t}\right) = -\int \left(\frac{\partial \rho}{\partial t}\log\rho\right) \tag{8}$$

where $\frac{\partial \rho}{\partial t}$ denotes partial derivative with respect to time. Now, noting the characterization of the Wasserstein norm from (7), we see that the steepest ascend direction (with respect to the Wasserstein metric) is given by $g = -\log\rho$. This gives

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla \log \rho) = \Delta \rho$$

which is the *linear heat equation*.

III. Matricial Wasserstein Geometry

For a range of problems in spectral analysis of vector-valued time series as well as in quantum mechanics, statistics of the underlying experimental setting are encapsulated in matrix-based models. For instance, in quantum mechanics, the statistical description of a system is via a state ρ, which is a positive element in a corresponding C^*-algebra of operators on a Hilbert space. For us, the Hilbert space will always be finite dimensional, and hence, ρ would simply be a positive-semidefinite Hermitian matrix with trace one. Likewise, in multivariable time series and vector-valued random variables (see, e.g., [4]), ρ may represent a matrix-valued power spectral density or a covariance matrix. In those cases, the integral of the trace or the trace, respectively, represent power and can be normalized 1 for our purposes.

Our aim is to develop a geometric framework that will have bearing on problems in quantum information theory as well as multivariable time series. Throughout the rest of this paper, the ρ s represent density matrices (positive-definite Hermitian matrices of trace one, or suitably normalized positive-definite Hermitian-valued functions), and we develop a noncommutative counterpart of the Wasserstein geometry by building on Quantum Mechanical insights and constructs. The key is to *devise a suitable notion of a continuity equation* as well as a *matrix-valued counterpart of the Benamou–Brenier action integral*. These are done next.

A. Quantum Continuity Equation

Our approach is based on the *Lindblad equation*, which describes the evolution of open quantum systems. These are thought of as coupled to a larger system (the environment, ancilla) and, thereby, cannot, in general, be described by a wave function. The proper description is in terms of a density operator ρ, which, in turn, obeys Lindblad's equation [6] (in diagonal form)

$$\dot\rho = -i[H,\rho] + \sum_{k=1}^{N}(L_k\rho L_k^* - \frac{1}{2}\rho L_k^* L_k - \frac{1}{2}L_k^* L_k \rho) \tag{9}$$

where * denotes the conjugate transpose, H is a Hamiltonian, and $[H,\rho] := H\rho - \rho H$ is the Lie bracket. The first term on the right-hand side describes the evolution of the state under the effect of the Hamiltonian H, and it is unitary (energy preserving). The other terms on the right-hand side model diffusion and, thereby, capture the dissipation of energy—it is the quantum analog of Laplace's operator Δ. The calculus we develop next highlights and actually underscores the parallels.

Regarding notation, we denote by $\mathcal{H}$ and $\mathcal{S}$ the set of $n \times n$ Hermitian and skew-Hermitian matrices, respectively. Since n is fixed throughout, we dispense of n in the notation. We also denote the space of block-column vectors consisting of N elements in $\mathcal{H}$ and $\mathcal{S}$ as $\mathcal{H}^N$ and $\mathcal{S}^N$, respectively. We let $\mathcal{H}_+$ and $\mathcal{H}_{++}$ denote the cones of nonnegative and positive-definite matrices, respectively, and

$$\mathcal{D}_+ := \{\rho \in \mathcal{H}_{++} \mid \operatorname{tr}(\rho) = 1\}. \tag{10}$$

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

$$T_\rho = \{\delta \in \mathcal{H} \mid \operatorname{tr}(\delta) = 0\}. \tag{11}$$

We use the standard notion of inner product

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

for both $\mathcal{H}$ and $\mathcal{S}$. For $X, Y \in \mathcal{H}^N$ ($\mathcal{S}^N$), we have

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

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

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

Throughout, we make the assumption that $L_k = L_k^*$, i.e., $L_k \in \mathcal{H}$ for all $k \in 1 \ldots, N$. Under this assumption, we define

$$\nabla_L : \mathcal{H} \to \mathcal{S}^N,\ X \mapsto \begin{bmatrix} L_1 X - X L_1 \\ \vdots \\ L_N X - X L_N \end{bmatrix} \tag{12}$$

as a *gradient operator*. Note that ∇_L acts just like the standard gradient operator and shares many useful properties such as[5]

$$\begin{aligned}\nabla_L(XY + YX) &= (\nabla_L X)Y + X(\nabla_L Y) \\ &\quad + (\nabla_L Y)X + Y(\nabla_L X)\ \forall X, Y \in \mathcal{H}.\end{aligned} \tag{13}$$

The dual of ∇_L, which is an analog of the (negative) *divergence operator*, is given by

$$\nabla_L^* : \mathcal{S}^N \to \mathcal{H},\ Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \mapsto \sum_{k}^{N} L_k Y_k - Y_k L_k. \tag{14}$$

Hence, the duality $\langle \nabla_L X, Y\rangle = \langle X, \nabla_L^* Y\rangle$ is straightforward.

With these definitions, we can easily calculate the (matricial) *Laplacian* as

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

which is exactly (after scaling by $1/2$) the diffusion term in the Lindblad equation (9). Therefore, the Lindblad equation (under the assumption that $L_k = L_k^*$) can be rewritten as

$$\dot\rho = -i[H,\rho] + \frac{1}{2}\Delta_L \rho.$$

Moreover, using the gradient operator (12) and its adjoint (14), we can now introduce a corresponding matricial *continuity equation* and, in fact, a family of such equations

$$\dot\rho = \nabla_L^* M_\rho(v) \tag{15}$$

[5]The domain of ∇_L is the space of Hermitian matrices; hence, we write $XY + YX$, instead of simply XY in (13).

where $M_\rho(v)$ can be any *noncommutative multiplication* between ρ and v that maps the "velocity field" $v \in \mathcal{S}^N$ to a "momentum field" $M_\rho(v) \in \mathcal{S}^N$.

Usually, in the Lindblad equation (9), 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 (15), 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 $L_1, \ldots, L_N$ to be a basis of the Hermitian matrices $\mathcal{H}$, in which case $N = n(n+1)/2$. Obviously, this construction ensures that the null space of ∇_L is spanned by I. Throughout this paper, the basis $L_1, \ldots, L_N$ is assumed to be fixed.

We will consider two interesting cases of noncommutative multiplication and the corresponding continuity equation, each of which has its own distinct properties.[6] The first case will be for

$$M_\rho(v) := \frac{1}{2}(v\rho + \rho v) \tag{16a}$$

which gives

$$\dot\rho = \frac{1}{2}\nabla_L^*(v\rho + \rho v) \tag{16b}$$

and $v = [v_1^*, \ldots, v_N^*]^* \in \mathcal{S}^N$. Clearly, $v\rho + \rho v \in \mathcal{S}^N$, which is consistent with the definition of ∇_L^*. We will refer to this as the *anticommutator* case, as it is standard to refer to

$$v\rho + \rho v =: \{\rho, v\}$$

as the anticommutator when applied to elements of an associative algebra. The second case will be for the Feynman–Kubo–Mori [18]–[20] product

$$M_\rho(v) := \int_0^1 \rho^s v \rho^{1-s} ds \tag{17a}$$

which leads to a continuity equation

$$\dot\rho = \nabla_L^* \int_0^1 \rho^s v \rho^{1-s} ds \tag{17b}$$

that we will refer to as the *logarithmic* case. This is the case that was extensively analyzed in [8] and [9] as alluded to in Section I. Here too, $\int_0^1 \rho^s v \rho^{1-s} ds \in \mathcal{S}^N$, which is consistent with the definition of ∇_L^*. The terminology "logarithmic" will become clearer in Section IV-B. The analysis of both equations and the resulting Wasserstein metrics is quite similar. Both give a fluid dynamic formulation of optimal transport on the space $\mathcal{D}_+$ of density matrices, thereby extending the work of Benamou and Brenier [5]. We will begin with the anticommutator case and then sketch the logarithmic one, both in the next section.

B. Matricial OMT

We treat separately the anticommutator and logarithmic cases of the two alternative noncommutative products $M_\rho(v)$ between ρ and v. The developments are completely analogous.

1) Anticommutator Formulation: Given two density matrices $\rho_0, \rho_1 \in \mathcal{D}_+$, one can formulate the optimization problem

$$W_{2,a}(\rho_0,\rho_1)^2 := \min_{\rho\in\mathcal{D}_+, v\in\mathcal{S}^N} \int_0^1 \operatorname{tr}(\rho v^* v) dt \tag{18a}$$

$$\dot\rho = \frac{1}{2}\nabla_L^*(v\rho + \rho v) \tag{18b}$$

$$\rho(0) = \rho_0,\ \rho(1) = \rho_1 \tag{18c}$$

[6]An interesting third case that is not discussed herein is $M_\rho(v) := \rho^{1/2} v \rho^{1/2}$.

and define the (matricial, "anticommutator") *Wasserstein distance* $W_{2,a}(\rho_0,\rho_1)$ between ρ_0 and ρ_1 to be the square root of the minimum of the cost (18a). More precisely, the minimum is taken over all the piecewise smooth paths connecting ρ_0 and ρ_1. Note here we have adopted the notation that $v^*v = \sum_{k=1}^N v_k^* v_k$ for $v \in \mathcal{S}^N$.

Let $\lambda(\cdot) \in \mathcal{H}$ be a smooth Lagrangian multiplier for the constraints (18b) and construct the Lagrangian

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

Pointwise minimizing the above over v yields

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

The corresponding minimum is

$$\begin{aligned}&\int_0^1 \left\{ -\frac{1}{2}\operatorname{tr}(\rho(\nabla_L\lambda)^*(\nabla_L\lambda)) - \operatorname{tr}(\dot\lambda\rho)\right\} dt \\ &\quad + \operatorname{tr}(\lambda(1)\rho_1) - \operatorname{tr}(\lambda(0)\rho_0)\end{aligned}$$

from which we conclude the following sufficient conditions for optimality. This optimality condition should be compared with (3).

Theorem 2: Suppose that there exist $\lambda(\cdot) \in \mathcal{H}, \rho(\cdot) \in \mathcal{D}_+$ satisfying

$$\dot\lambda + \frac{1}{2}(\nabla_L\lambda)^*(\nabla_L\lambda) = 0 \tag{19a}$$

$$\dot\rho - \frac{1}{2}\nabla_L^*(\nabla_L\lambda\,\rho + \rho\nabla_L\lambda) = 0 \tag{19b}$$

together with the boundary conditions $\rho(0) = \rho_0, \rho(1) = \rho_1$; then, the pair (ρ, v) with $v = \nabla_L\lambda$ solves (18).

The Wasserstein distance $W_{2,a}$ gives a Riemannian structure

$$\langle \delta_1, \delta_2\rangle_\rho = \frac{1}{2}\operatorname{tr}(\rho\nabla\lambda_1^*\nabla\lambda_2 + \rho\nabla\lambda_2^*\nabla\lambda_1)$$

on the tangent space (11). Here, λ_j, $j = 1, 2$, is the solution to the "Poisson" equation

$$\delta_j = \frac{1}{2}\nabla_L^*(\nabla_L\lambda_j\,\rho + \rho\nabla_L\lambda_j). \tag{20}$$

The proof of existence and uniqueness of the solution of (20) follows exactly along the same lines as in [21, Sec. 3.2]. See also [22] for details. In fact, given a tangent vector δ, $\nabla_L\lambda$ is the unique minimizer of $\operatorname{tr}(\rho v^* v)$ over all the velocities $v \in \mathcal{S}^N$ satisfying

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

Therefore, with the above definition of Riemannian structure, $W_{2,a}(\cdot,\cdot)$ indeed defines a metric on $\mathcal{D}_+$. Moreover, the distance between two given $\rho_0, \rho_1 \in \mathcal{D}_+$ can be rewritten as

$$W_{2,a}(\rho_0,\rho_1) = \min_\rho \int_0^1 \sqrt{\langle\dot\rho(t), \dot\rho(t)\rangle_{\rho(t)}}\, dt$$

where the minimum is taken over all the piecewise smooth paths on the manifold $\mathcal{D}_+$.

The Wasserstein distance $W_{2,a}$ can be extended to the closure of $\mathcal{D}_+$, i.e., the space (denoted by $\mathcal{D}$) of all positive-semidefinite matrices with trace 1, by continuity. For any two matrices $\rho_0, \rho_1 \in \mathcal{D}$, we can construct sequences $\{\rho_0^j\}, \{\rho_1^j\}$ in $\mathcal{D}_+$ converging to ρ_0 and ρ_1, respectively, in the Frobenius norm. It can be shown that the definition $W_2(\rho_0,\rho_1) := \lim_{j\to\infty} W_2(\rho_0^j, \rho_1^j)$ makes sense; see [21, Proposition 4.5].

Remark 1: For computational purposes, it is important to note that Problem (18) can be cast as a convex optimization problem in a manner analogous to that in the scalar case [5]; cf., (4). Define $u := v\rho = [u_1^*, \dots, u_N^*]^*$ and $\bar{u} := [u_1, \dots, u_N]^*$; then

$$\begin{aligned}\operatorname{tr}(\rho v^* v) &= \sum_{k=1}^{N} \operatorname{tr}(\rho v_k^* v_k) \\ &= \sum_{k=1}^{N} \operatorname{tr}(v_k \rho \rho^{-1} (v_k \rho)^*) = \operatorname{tr}(u \rho^{-1} u^*)\end{aligned}$$

and we readily arrive at the equivalent convex optimization problem

$$W_{2,a}(\rho_0, \rho_1)^2 = \min_{\rho, u} \int_0^1 \operatorname{tr}(u \rho^{-1} u^*) dt \tag{21a}$$

$$\dot{\rho} = \frac{1}{2} \nabla_L^* (u - \bar{u}) \tag{21b}$$

$$\rho(0) = \rho_0, \ \rho(1) = \rho_1. \tag{21c}$$

The minimum is taken over all the piecewise smooth functions ρ and u. Even though we do not put any structure constraint on u, the optimal u satisfies $u = v\rho$ for some $v \in \mathcal{S}^N$ [22]. An efficient algorithm is presented in [23].

2) Transport With Spatial Component: In applications, it is often the case that one has to deal with matrix-valued distributions along spatial or frequency axes. Thus, in this case, the ρ s may be $\mathcal{H}_+$-valued functions on a subdomain $E \subset \mathbb{R}^m$. For instance, in the context of multivariable time-series analysis, it is natural to consider $m = 1$, and $E = [0, \pi]$; see, e.g., [4]. For simplicity, we assume E to be a (convex) connected compact set. Therefore, in this section

$$\mathcal{D} = \{\rho(\cdot) \mid \rho(x) \in \mathcal{H}_+ \text{ for } x \in E \text{ such that } \int_E \operatorname{tr}(\rho(x)) dx = 1\}.$$

Let $\mathcal{D}_+$ denote the interior of $\mathcal{D}$, and in order to avoid proliferation of notation, we use the same symbol $\mathcal{D}$ ($\mathcal{D}_+$) as above. By combining the standard continuity equation on the Euclidean space and the continuity equation for density matrices (16b), we obtain a continuity equation on $\mathcal{D}_+$ for the flow $\rho(t, x)$ as

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w\rho + \rho w) - \frac{1}{2} \nabla_L^* (v\rho + \rho v) = 0. \tag{22}$$

Here, $\nabla_x \cdot$ is the standard (negative) divergence operator on $\mathbb{R}^m$, $w(t, x) \in \mathcal{H}^m$ is the velocity field along the space dimension, and $v(t, x) \in \mathcal{S}^N$ is the quantum velocity as before.

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

$$W_{2,a}(\rho_0, \rho_1)^2 := \min \int_0^1 \int_E \{\operatorname{tr}(\rho w^* w) + \gamma \operatorname{tr}(\rho v^* v)\} dx dt \tag{23a}$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (w\rho + \rho w) - \frac{1}{2} \nabla_L^* (v\rho + \rho v) = 0 \tag{23b}$$

$$\rho(0, \cdot) = \rho_0, \ \rho(1, \cdot) = \rho_1. \tag{23c}$$

The minimum is taken over piecewise smooth functions $\rho(\cdot) \in \mathcal{D}_+, w(\cdot) \in \mathcal{H}^m, v(\cdot) \in \mathcal{S}^N$ under zero-flux condition on the boundary ∂E. The coefficient $\gamma > 0$ is arbitrary and weighs in the relative significance of the two velocity fields. It is anticipated that, in applications, a suitable choice of γ will provide appropriate flows that reflect the underlying physics (trading off the two alternative mechanisms for transferring mass, i.e., via "flow along x" or via the available "noncommutative flow"). Once again, we are in a position to define a Wasserstein distance $W_{2,a}(\rho_0, \rho_1)$ between ρ_0 and ρ_1 via (23a).

A sufficient condition for optimality can be obtained in a similar manner as before. Here, we let $\lambda(\cdot, \cdot) \in \mathcal{H}$ be a smooth function and define the Lagrangian

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

Integration by parts yields

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

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

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

and

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

Substituting these back to the Lagrangian, we obtain

$$\begin{aligned}\int_0^1 \int_E \Big\{ &-\frac{1}{2} \operatorname{tr}(\rho (\nabla_x \lambda)^* (\nabla_x \lambda)) \\ &- \frac{1}{2\gamma} \operatorname{tr}(\rho (\nabla_L \lambda)^* (\nabla_L \lambda)) - \operatorname{tr}\Big(\rho \frac{\partial \lambda}{\partial t}\Big)\Big\} dx dt\end{aligned}$$

and the sufficient conditions for optimality given below follow.

Theorem 3: Suppose that there exist smooth $\lambda(\cdot, \cdot) \in \mathcal{H}, \rho(\cdot, \cdot) \in \mathcal{D}_+$ satisfying

$$\frac{\partial \lambda}{\partial t} + \frac{1}{2} (\nabla_x \lambda)^* (\nabla_x \lambda) + \frac{1}{2\gamma} (\nabla_L \lambda)^* (\nabla_L \lambda) = 0 \tag{24a}$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2} \nabla_x \cdot (\nabla_x \lambda \, \rho + \rho \nabla_x \lambda) - \frac{1}{2\gamma} \nabla_L^* (\nabla_L \lambda \, \rho + \rho \nabla_L \lambda) = 0 \tag{24b}$$

together with boundary conditions $\rho(0, \cdot) = \rho_0, \rho(1, \cdot) = \rho_1$; then, $(\rho, w, v) = (\rho, \nabla_x \lambda, \frac{1}{\gamma} \nabla_L \lambda)$ solves (23).

The Wasserstein distance $W_{2,a}$ defines a Riemannian-type structure on the tangent space of $\mathcal{D}_+$ at ρ. Given any two tangent vectors δ_1, δ_2 at ρ, we can associate them with λ_1, λ_2 by solving the "Poisson" equations

$$\delta_j = \frac{1}{2\gamma} \nabla_L^* (\nabla_L \lambda_j \, \rho + \rho \nabla_L \lambda_j) - \frac{1}{2} \nabla_x \cdot (\nabla_x \lambda_j \, \rho + \rho \nabla_x \lambda_j) \tag{25}$$

for $j = 1, 2$. Similar to the argument we had before the case (20) without a spatial component, the above Poisson equation (25) has an unique solution. The proof relies on the fact that the null space of the gradient operator

$$\nabla_{L,x} = \begin{bmatrix} \nabla_L \\ \nabla_x \end{bmatrix}$$

is spanned by the constant matrix function I.

The Riemannian metric can then be defined as

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

Therefore, $W_{2,a}(\cdot,\cdot)$ is a metric on $\mathcal{D}_+$ and can be rewritten as

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

Here, the integral is minimized over all the piecewise smooth curves in $\mathcal{D}_+$ connecting ρ_0 and ρ_1. As in Section III-B1, the Wasserstein distance $W_{2,a}$ can be extended to the closure $\mathcal{D}$ of $\mathcal{D}_+$ by continuity.

Remark 2: As noted earlier, (23) can again be cast as a convex optimization problem: define $q = w\rho$ to obtain the equivalent form

$$\min_{\rho,q,u} \int_0^1 \int_E \left\{ \operatorname{tr}(q\rho^{-1}q^*) + \gamma \operatorname{tr}(u\rho^{-1}u^*) \right\} dx dt \tag{26a}$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{2}\nabla_x \cdot (q + \bar{q}) - \frac{1}{2}\nabla_L^*(u - \bar{u}) = 0 \tag{26b}$$

$$\rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1. \tag{26c}$$

3) Logarithmic Case: We now briefly discuss the case where the noncommutative multiplication of ρ and v is taken to be (17a):

$$M_\rho(v) = \int_0^1 \rho^s v \rho^{1-s} ds.$$

For the purposes of defining a corresponding Wasserstein geometry, we proceed in a manner entirely analogous to that for the anticommutator case. Hence, we only highlight the key elements.

The corresponding Wasserstein metric between $\rho_0, \rho_1 \in \mathcal{D}_+$ is obtained via

$$W_{2,b}(\rho_0,\rho_1)^2 := \min_{\rho\in\mathcal{D}_+, v\in\mathcal{S}^N} \int_0^1 \int_0^1 \operatorname{tr}(v^*\rho^s v \rho^{1-s}) ds dt \tag{27a}$$

$$\dot{\rho} = \nabla_L^* \int_0^1 \rho^s v \rho^{1-s} ds \tag{27b}$$

$$\rho(0) = \rho_0,\ \rho(1) = \rho_1. \tag{27c}$$

Employing a similar argument as in Theorem 2 (see also [21, Th. 5.3]), we establish the following optimality condition.

Proposition 4: Suppose that there exist $\lambda(\cdot) \in \mathcal{H}, \rho(\cdot) \in \mathcal{D}_+$ satisfying

$$\dot{\lambda} = -\int_0^1 \int_0^1 \int_0^\alpha \left\{ \frac{\rho^{\alpha-\beta}}{(1-s)I + s\rho} (\nabla_L \lambda)^* \rho^{1-\alpha} \times \nabla_L \lambda \frac{\rho^\beta}{(1-s)I + s\rho} \right\} d\beta d\alpha ds \tag{28a}$$

$$\dot{\rho} = \nabla_L^* \int_0^1 \rho^s \nabla_L \lambda \rho^{1-s} ds \tag{28b}$$

together with the boundary conditions $\rho(0) = \rho_0, \rho(1) = \rho_1$; then, the pair (ρ, v) with $v = \nabla_L \lambda$ solves (27).

The above optimality condition should be compared with (3) in the scalar case and Theorem 2 in the anticommutator case. Unlike the other two, where ρ does not affect λ directly; here, the two differential equations (28) are coupled in both directions.

Similarly, for matrix-valued densities, the corresponding metric is obtained via

$$\min_{\rho,w,v} \int_0^1 \int_E \int_0^1 \left\{ \operatorname{tr}(w^*\rho^s w \rho^{1-s}) + \gamma \operatorname{tr}(v^*\rho^s v \rho^{1-s}) \right\} ds dx dt$$

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot \left(\int_0^1 \rho^s w \rho^{1-s} ds \right) - \nabla_L^* \left(\int_0^1 \rho^s v \rho^{1-s} ds \right) = 0$$

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

IV. Gradient Flow of the Entropy

We close by presenting the matricial counterpart of the classical result of [14] for the case of scalar-valued distributions that the gradient flow of the entropy is the heat equation; see Section II-C. Thus, in the following, we derive gradient flows for the entropy functional on density matrices with respect to the two alternative Wasserstein geometries.

A. Anticommutator Case

The entropy of density matrices is defined by

$$S(\rho) = -\operatorname{tr}(\rho \log \rho).$$

The gradient with respect to $W_{2,a}$ may be calculated as follows. For a given flow $\rho(\cdot)$, we have

$$\begin{aligned} \frac{dS(\rho(t))}{dt} &= -\operatorname{tr}((\log\rho + I)\dot{\rho}) \\ &= -\operatorname{tr}((\log\rho + I)\frac{1}{2}\nabla_L^*(v\rho + \rho v)) \\ &= -\frac{1}{2}\operatorname{tr}((\nabla_L \log\rho)^*(v\rho + \rho v)) \\ &= -\frac{1}{2}\operatorname{tr}(\rho v^* \nabla_L \log\rho + \rho(\nabla_L \log\rho)^* v). \end{aligned}$$

In view of the definition of $W_{2,a}$, we conclude that the steepest ascent direction is

$$v = -\nabla_L \log \rho.$$

Substituting back to the continuity equation (16b), we obtain the gradient flow

$$\dot{\rho} = -\frac{1}{2}\nabla_L^*(\nabla_L \log\rho\,\rho + \rho\nabla_L \log\rho) = -\frac{1}{2}\nabla_L^*\{\rho, \nabla_L \log\rho\} \tag{29}$$

where $\{\cdot,\cdot\}$ denotes the anticommutator as before.

Similarly, we may consider the entropy function for matrix-valued densities

$$S(\rho) = -\int_{\mathbb{R}^m} \operatorname{tr}(\rho \log \rho) dx$$

and the associated gradient flow with respect to $W_{2,a}$. The total derivative of S over a flow $\rho(t,\cdot)$ is

$$\begin{aligned} \frac{dS(\rho(t,\cdot))}{dt} &= -\int_{\mathbb{R}^m} \operatorname{tr}((\log\rho + I)\frac{\partial\rho}{\partial t}) dx \\ &= \int_{\mathbb{R}^m} \left\{ -\frac{1}{2}\operatorname{tr}(\rho w^* \nabla_x \log\rho + \rho(\nabla_x \log\rho)^* w) \right. \\ &\quad \left. -\frac{1}{2}\operatorname{tr}(\rho v^* \nabla_L \log\rho + \rho(\nabla_L \log\rho)^* v) \right\} dx \end{aligned}$$

which indicates, in view of (23), that the steepest ascent direction is

$$w = -\nabla_x \log\rho,\ v = -\frac{1}{\gamma}\nabla_L \log\rho.$$

Therefore, the gradient flow is now given by

$$\frac{\partial\rho}{\partial t} = \frac{1}{2}\nabla_x \cdot \{\rho, \nabla_x \log\rho\} - \frac{1}{2\gamma}\nabla_L\{\rho, \nabla_L \log\rho\}.$$

Remark 3: Note that in both of the above cases, the gradient flow of the entropy is nonlinear, which should be contrasted with the linear heat equation that arises in the scalar case (as noted in Section II-C following [14] and [15]). Indeed, (29) is a second-order nonlinear equation, which is quite different from the linear Lindblad equation, and gives the direction of maximal dissipation of quantum information relative to the Wasserstein metric $W_{2,a}$ defined above.

B. Logarithmic Case

In this section, we will see that in the logarithmic case, i.e., when using the noncommutative multiplication and corresponding continuity equation in (17), the gradient flow with respect to the corresponding Wasserstein geometry of the matricial entropy now gives the quantum version of the heat equation $\dot{\rho} = \Delta_L \rho$, which is the "dissipative part" of the Lindblad equation. The linear heat equation for the gradient flow of entropy in the logarithmic case was independently derived in [8] and [9].

A key property of, and our choice of the terminology "logarithmic" for, the noncommutative multiplication (17a) can be traced in the rather *remarkable identity* (see [8] for details)

$$\nabla_L \rho = \int_0^1 \rho^s (\nabla_L \log \rho) \rho^{1-s} ds. \tag{30}$$

It represents a logarithmic averaging, although, at first surprising, the identity itself may be readily proven using the product rule [see (13)] and the fact that

$$\rho = \lim_{j\to\infty} (I + \frac{1}{j} \log \rho)^j.$$

The relation (30) works just as well for general gradient operators

$$\nabla_x \rho = \int_0^1 \rho^s (\nabla_x \log \rho) \rho^{1-s} ds. \tag{31}$$

With this in mind, we move onto the gradient flow of the entropy $S(\rho)$ with respect to $W_{2,b}$. For the case where $\rho(t) \in \mathcal{D}_+$, i.e., ρ does not depend on spatial coordinates, taking the total derivative of $S(\rho)$ over a flow $\rho(\cdot)$ gives

$$\begin{aligned} \frac{dS(\rho(t))}{dt} &= -\operatorname{tr}((\log \rho + I)\dot{\rho}) \\ &= -\operatorname{tr}((\nabla_L \log \rho)^* \int_0^1 \rho^s v \rho^{1-s} ds) \end{aligned}$$

which points to the greatest ascent direction $v = -\nabla_L \log \rho$. Now, using (30), we obtain

$$\dot{\rho} = -\nabla_L^* \int_0^1 \rho^s (\nabla_L \log \rho) \rho^{1-s} ds = -\nabla_L^* \nabla_L \rho = \Delta_L \rho \tag{32}$$

which is a *linear heat equation*, just as in the scalar case.

Similarly, employing (31), we see (arguing as above) that the gradient flow of $S(\rho)$ with respect to $W_{2,b}$ with spatial dimension is

$$\frac{\partial \rho}{\partial t} = \Delta_L \rho + \Delta_x \rho. \tag{33}$$

Remark 4: First and foremost, (32) and (33) are indeed quite intriguing because of their similarity to the scalar case. But, more importantly, (32) implies that the dissipation part of the Lindblad equation gives a direction in which quantum information (negative of quantum entropy) is decreasing as rapidly as possible with respect to the specific Wasserstein geometry ($W_{2,b}$) on the space of density matrices.

Remark 5: It should be noted that both heat equations (29) and (32) can be written in the form

$$\dot{\rho} = -\nabla_L^* M_\rho (\nabla_L \log \rho) \tag{34}$$

but with different noncommutative multiplications. This formula of gradient flow of the entropy S even holds for the cases of other more general noncommutative multiplications. The fact that the heat equation becomes linear in the logarithmic case is due to the remarkable relation (30).

V. Example

We now highlight the relevance of our results with an illustrative example. In this, we use the matricial Wasserstein-2 to compare and interpolate two multivariate power spectra. Thus, the data of the example consist of two 2×2 matrix-valued densities, specified at $t = 0$ and $t = 1$, respectively, and distributed over a frequency interval $[0, \pi]$. Fig. 1 depicts the interpolation of these two distributions using (23). The entry $\rho(1,1)$ represents the energy density in the first channel, while $\rho(2,2)$ represents the energy density in the second. The figure exemplifies how the energy transfers gradually from channel 1 to channel 2, while, at the same time, the dominant frequency mode shifts smoothly between the respective two peak frequencies. This is a characteristic of our matricial-OMT interpolation. Evidently, linear interpolation of the two end-point spectra fails to show such a shift in power and frequency, which is natural and highly desirable. Indeed, linear interpolation is plagued by the push-pop effect, where the peak power at one frequency reduces, while power appears and increases at another frequency at the same time. Push-pop effects are inconsistent with the underlying physics, where energy density shifts with change in position of the underlying cause.

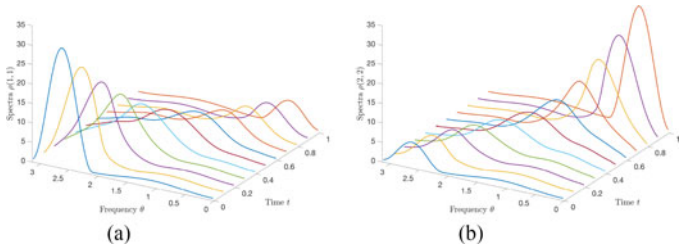

Fig. 1. Interpolation of power spectra. (a) $\rho(1,1)$. (b) $\rho(2,2)$.

The intent of such interpolation is to be used in modeling and morphing time-varying signals and power distributions. Also, since the dynamics of multi-input multi-output (MIMO) systems are closely related to the power spectrum of the input–output process when driven by noise, the approach is expected to be relevant comparing, approximating, and interpolating MIMO system dynamics as well.

VI. Conclusion and Further Research

In this paper, we proposed a possible extension of the Benamou–Brenier control-theoretic approach to OMT to the noncommutative case of probability density matrices using ideas from quantum theory. We discussed two cases where the noncommutative multiplication is taken to be the anticommutator and logarithmic multiplication between density matrices and matricial velocity fields, respectively. Each of them has certain advantages relative to one another. The anticommutator case can be formulated as a convex optimization problem, where the optimality condition resembles the one in the scalar setting, while the logarithmic mean case leads to a linear heat equation as the gradient flow of the entropy.

An immediate application of the Wasserstein distance that we define to matrix-valued distribution as in [4] provides a computable metric to compare matrix-valued power spectral densities and, thereby, compare multivariable time series. However, the need to compare matrix-valued distributions in a quantitative manner is evident in many other areas of current interest in the field of controls, such as image-based analysis, information and quantum control, and the study of networks.

Specifically, in medical image processing, a key technique is DTI, which is used extensively in white matter tractography [24]. Here, one assigns a tensor to each point in space in order to track the diffusion of water molecules in neuronal fibers and, thus, image the white matter tracts of the brain. Such imagery is used in surgical planning as well as in studying various neurological disorders such as Parkinson's, Alzheimer's disease, and schizophrenia. Understanding the white matter connections allows us to visualize the complex circuitry of the brain. Thus, the geometry and distances described in this paper are directly applicable to the type of matrix-valued fields encountered in DTI and promises a potentially important research direction where tracking of

neuronal fiber information could be approached via suitable matrix-flow control problems as suggested herein.

On the topic networks and network dynamics, it is interesting to point out that "*network robustness*" is intimately related to entropy [25] and thus via the work of [26] and [27] to curvature. Network robustness is also closely related to feedback redundancy for systems modeled by weighted graphs with important implications for analyzing the mechanisms of resistance in cancer [28], [29]. The conceptual basis of robustness of networks, at the scales that one needs to consider, is intimately related to OMT via Ollivier–Ricci curvature, as discussed in these references.

It is, thus, natural from a control as well as an applications perspective to consider generalization of transport problems to vector-valued as well as matrix-valued quantities across relevant underlying spaces, e.g., networks, and thereby suitable generalization of curvature along the matrix-valued framework of the current paper. We feel that this is a promising direction of research. Finally, we expect that (34) will have applications to quantum information theory and quantum networks. The gradient flows proposed in this paper identify directions in which entropy is increasing in the steepest manner (relative to the given Wasserstein distance) and consequently in which information is decreasing the fastest.

Acknowledgment

The authors would like to thank Dr. K. Yamamoto, Dr. L. Ning, and Dr. S. Z. Khong for some very helpful conversations about matrix OMT.

References

[1] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*. New York, NY, USA: Springer, 1998.
[2] C. Villani, *Topics in Optimal Transportation*. Providence, RI, USA: Amer. Math. Soc., 2003.
[3] L. V. Kantorovich, "On a problem of Monge," *J. Math. Sci.*, vol. 3, pp. 225–226, 1948.
[4] L. Ning, T. T. Georgiou, and A. Tannenbaum, "On matrix-valued Monge-Kantorovich optimal mass transport," *IEEE Trans. Automat. Control*, vol. 60, no. 2, pp. 373–382, Feb. 2015.
[5] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem," *Numer. Math.*, vol. 84, no. 3, pp. 375–393, 2000.
[6] S. J. Gustafson and I. M. Sigal, *Mathematical Concepts of Quantum Mechanics*. New York, NY, USA: Springer, 2011.
[7] Y. Chen, T. T. Georgiou, and A. Tannenbaum, "Matrix optimal mass transport: A quantum mechanical approach," *arXiv preprint arXiv:1610.03041*, 2016.
[8] E. A. Carlen and J. Maas, "Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance," *arXiv preprint arXiv:1609.01254*, 2016.
[9] M. Mittnenzweig and A. Mielke, "An entropic gradient structure for Lindblad equations and generic for quantum systems coupled to macroscopic models," *arXiv preprint arXiv:1609.05765*, 2016.
[10] Y. Chen, T. T. Georgiou, and A. Tannenbaum, "Interpolation of density matrices and matrix-valued measures: The unbalanced case," *arXiv preprint arXiv:1612.05914*, 2016.
[11] Y. Chen, T. T. Georgiou, L. Ning, and A. Tannenbaum, "Matricial Wasserstein-1 distance," *IEEE Control Syst. Lett.*, vol. 1, no. 1, pp. 14–19, Jul. 2017.
[12] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part I," *IEEE Trans. Automat. Control*, vol. 61, no. 5, pp. 1158–1169, May 2016.
[13] Y. Chen, T. T. Georgiou, and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint," *J. Optim. Theory Appl.*, vol. 169, no. 2, pp. 671–691, 2016.
[14] F. Otto, "The geometry of dissipative evolution equations: The porous medium equation," *Commun. Partial Differential Equations*, vol. 26, pp. 101–174, 2001.
[15] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation," *SIAM J. Math. Anal.*, vol. 29, no. 1, pp. 1–17, 1998.
[16] E. Tannenbaum, T. Georgiou, and A. Tannenbaum, "Signals and control aspects of optimal mass transport and the Boltzmann entropy," in *Proc. 49th IEEE Conf. Decision Control*, 2010, pp. 1885–1890.
[17] R. J. McCann , "A convexity principle for interacting gases," *Adv. Math.*, vol. 128, no. 1, pp. 153–179, 1997.
[18] R. P. Feynman, "An operator calculus having applications in quantum electrodynamics," *Phys. Rev.*, vol. 84, no. 1, 1951, Art. no. 108.
[19] F. Hiai, D. Petz, and G. Toth, "Curvature in the geometry of canonical correlation," *Stud. Sci. Math. Hungarica*, vol. 32, no. 1, pp. 235–250, 1996.
[20] M. Hayashi, *Quantum Information*. New York, NY, USA: Springer, 2006.
[21] E. A. Carlen and J. Maas, "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.*, vol. 331, no. 3, pp. 887–926, 2014.
[22] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum, "On the matrix Monge-Kantorovich problem," *arXiv preprint arXiv: 1701.02826*, 2017.
[23] Y. Chen, E. Haber, K. Yamamoto, T. T. Georgiou, and A. Tannenbaum, "An efficient algorithm for matrix-valued and vector-valued optimal mass transport," *arXiv preprint arXiv: 1706.08841*, 2017.
[24] W. Van Hecke, L. Emsell, and S. Sunaert, *Diffusion Tensor Imaging: A Practical Handbook*. New York, NY, USA: Springer, 2015.
[25] L. A. Demetrius, "Boltzmann, Darwin and directionality theory," *Phys. Rep.*, vol. 530, no. 1, pp. 1–85, 2013.
[26] J. Lott and C. Villani, "Ricci curvature for metric-measure spaces via optimal transport," *Ann Math.*, vol. 169, pp. 903–991, 2009.
[27] M.-K. von Renesse and K.-T. Sturm, "Transport inequalities, gradient estimates, entropy and Ricci curvature," *Commun. Pure Appl. Math.*, vol. 58, no. 7, pp. 923–940, 2005.
[28] R. Sandhu *et al.*, "Graph curvature for differentiating cancer networks," *Sci. Rep.*, vol. 5, 2015, Art. no. 12323.
[29] A. Tannenbaum *et al.*, "Graph curvature and the robustness of cancer networks," *arXiv preprint 1502.04512*, 2016.