HHS Public Access

Author manuscript

IEEE Control Syst Lett. Author manuscript; available in PMC 2018 July 01.

Published in final edited form as:

IEEE Control Syst Lett. 2017 July ; 1(1): 14–19. doi:10.1109/LCSYS.2017.2699319.

Matricial Wasserstein-1 Distance

Yongxin Chen,
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY

Tryphon T. Georgiou,
Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA

Lipeng Ning, and
Brigham and Women's Hospital (Harvard Medical School), MA

Allen Tannenbaum
Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY

Abstract

We propose an extension of the Wasserstein 1-metric (W_1) for density matrices, matrix-valued density measures, and an unbalanced interpretation of mass transport. We use duality theory and, in particular, a "dual of the dual" formulation of W_1. This matrix analogue of the Earth Mover's Distance has several attractive features including ease of computation.

I. Introduction

Optimal mass transport (OMT) has proven to be a powerful methodology for numerous problems in physics, probability, information theory, fluid mechanics, econometrics, systems and control, computer vision, and signal/image processing [1], [2], [3], [4], [5]. Developments along purely controls-related issues ensued when it was recognized that mass transport may be naturally reformulated as a stochastic control problem; see [6], [7], [8], [9], [10], [11], [12], [13], [14] and the references therein.

Historically, the problem of OMT [1], [3] began with the question of minimizing the effort of transporting one distribution to another with a *cost proportional to the Euclidean distance*, as in Monge's original formulation, between starting and ending points of the mass being transported. However, the control-theoretic reformulation [2] which was at the root of the aforementioned developments was based on the *choice of a quadratic cost*. The quadratic cost allowed the interpretation of the transport effort as an action integral and gave rise to a Riemannian structure on the space of distributions [15], [16], [17]. The originality in our present work is two-fold. First, we formulate the transport problem with an L_1 cost in a similar manner, as a control problem with an L_1-path cost functional, and secondly, we develop theory for shaping flows of matrixvalued distributions which is a non-trivial generalization of classical OMT.

The relevance of OMT on flows of matrix-valued distributions was already recognized in [18], [19] and was cast as a control problem as well, albeit in a quadratic-cost setting. At that

point, interest in the geometry of matrix-valued distributions stemmed from applications to spectral analysis of vectorvalued time series (see [18] and the references therein). Yet soon it became apparent that flows of matrix-valued distributions represent evolution of quantum systems. In fact, there has been a burst of activity in applying ideas of quantum mechanics to OMT of matrix-valued densities as well as, utilizing an OMT framework to study the dynamics of quantum systems: three groups [20], [21], [22] independently and simultaneously developed quantum mechanical frameworks for defining a Wasserstein-2 distance on density matrices, via a variational formalism generalizing the work of [2]. We note that [20], [21], [22] develop matricial generalizations of the Wasserstein 2-metric (W_2) and explore the *Riemannian-like* structure for studying the entropic flows of quantum states. While [20], [22] aimed to explain the Lindblad equation as gradient flow with respect to Wasserstein-2 metric, our motivation in [21] was to develop a transport distance for matrix-valued densities that is computable and suitable for engineering applications.

Thus, in our present note, along the lines of [21], we develop a natural extension of the Wasserstein 1-metric to density matrices and matrix-valued measures. Our point of view is somewhat different from the earlier works on matricial Wassserstein-2 metrics. We mainly use duality theory [23], [3]. Further, we do not employ the Benamou and Brenier [2] control formulation of OMT, but rather the Kantorovich-Rubinstein duality. This new scheme is computationally more attractive and, moreover, it is especially appealing when specialized to weighted graphs (discrete spaces) that are sparse (few edges), as is the case for many real-world networks [24], [25], [26].

II. Optimal mass transport

We begin with duality theory, explained for scalar densities, upon which our matricial generalization of the Wasserstein-1 metric is based.

Given two probability densities ρ_0 and ρ_1 on $\mathbb{R}^m$, the Wasserstein-1 distance between them is

$$W_1(\rho_0,\rho_1):=\inf_{\pi\in\prod(\rho_0,\rho_1)}\int_{\mathbb{R}^m\times\mathbb{R}^m}\|x-y\|\pi(\mathrm{dxdy}), \tag{1}$$

where $\|\cdot\|$ represents the Euclidean distance and $\prod(\rho_0,\rho_1)$ denotes the set of couplings between ρ_0 and ρ_1, i.e., the set of joint distributions of ρ_0 and ρ_1. The Wasserstein-1 distance has a dual formulation via the following result due to Kantorovich and Rubinstein [23], [1], [3]:

$$W_1(\rho_0,\rho_1)=\sup_f\left\{\int_{\mathbb{R}^m}f(x)(\rho_1(x)-\rho_0(x))dx\ \text{subject to}\|f\|_{\mathrm{Lip}}\le 1\right\}, \tag{2}$$

where $\|f\|_{\mathrm{Lip}}$ denotes the Lipschitz constant. When f is differentiable, $\|f\|_{\mathrm{Lip}}=\|\nabla_x f\|$. It follows that

$$W_1(\rho_0,\rho_1)=\sup_{f\in C^1}\left\{\int_{\mathbb{R}^m} f(x)(\rho_1(x)-(\rho_0(x))dx \text{ subject to } \|\nabla_x f\|\leq 1\right\}. \quad (3)$$

Starting from (3), by once again considering the dual, we readily obtain the very important reformulation [23]

$$W_1(\rho_0,\rho_1)=\inf_u\left\{\int_{\mathbb{R}^m}\|u(x)\|dx \text{ subject to } \rho_1-\rho_0+\nabla_x\cdot u=0\right\}, \quad (4)$$

where the optimization variable $u\in\mathbb{R}^m$ now becomes flux. Alternatively, this can be written as the control-optimization problem in the Benamou-Brenier style [2], [3]

$$W_1(\rho_0,\rho_1)=\inf_u\left\{\int_0^1\int_{\mathbb{R}^m}\|u(t,x)\|\text{dxdt subject to } \frac{\partial\rho(t,x)}{\partial t}+\nabla_x\cdot u(t,x)=0, \text{ and } \rho(0,x)=\rho_0(x),\rho_1(1,x)=\rho_1(x)\right\}, \quad (5)$$

This “dual of the dual” formulation turns the Kantorovich and Rubinstein into a control problem to determine a suitable velocity (control vector) *u*.

Remark 1

We remark that from a computational standpoint, when applied to discrete spaces (graphs), this formulation leads to a very substantial computational benefit in the case of sparse graphs; this is due to the fact that (1) involves solving systems of the order of the square of the number of nodes, while equation (4), solving systems of the order of the number of edges. This fact has been utilized recently in studying biological networks [27]. See also [28], [29] for an efficient algorithm based on (4).

III. Gradient on space of Hermitian matrices

We closely follow the treatment in [21]. In particular, we will need a notion of gradient on the space of Hermitian matrices and its dual, i.e. the divergence.

Denote by $\mathscr{H}$, $\mathscr{S}$ and $\mathscr{C}$ the set of $n\times n$ Hermitian, skew-Hermitian and general complex-valued matrices, respectively. We will assume that all of our matrices are of fixed size $n\times n$. Next, denote the space of block-column vectors consisting of N elements in $\mathscr{S}$, $\mathscr{H}$ and $\mathscr{C}$ as $\mathscr{S}^N$, $\mathscr{H}^N$ and $\mathscr{C}^N$, respectively. We also let $\mathscr{H}_+$ and $\mathscr{H}_{++}$ denote the cones of nonnegative and positive-definite matrices, respectively, and

$$\mathscr{D}:=\{\rho\in\mathscr{H}_+|\text{tr}(\rho)=1\}, \quad (6)$$

$$\mathscr{D}_+ := \{\rho \in \mathscr{H}_{++} | \mathrm{tr}(\rho)=1\}. \quad (7)$$

The tangent space of $\mathscr{D}_+$, at any $\rho \in \mathscr{D}_+$ is given by

$$T_\rho = \{\delta \in \mathscr{H} | \mathrm{tr}(\delta)=0\}, \quad (8)$$

and we use the standard notion of inner product, namely

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

for $\mathscr{H}$, $\mathscr{S}$ and $\mathscr{C}$. For $X, Y \in \mathscr{H}^N(\mathscr{S}^N, \mathscr{C}^N)$,

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

Given $X=[X_1^*, \cdots, X_N^*]^* \in \mathscr{C}^N$, we denote $X_*=[X_1, \cdots, X_N]^* \in \mathscr{C}^N$ as the block-wise conjugate of X.

For a given $L \in \mathscr{H}^N$ we define

$$\nabla_L : \mathscr{H} \to \mathscr{S}^N, X \mapsto \begin{bmatrix} L_1 X & - & X L_1 \\ & \vdots & \\ L_N X & - & X L_N \end{bmatrix} \quad (9)$$

to be the ***gradient operator***. By analogy with the ordinary multivariable calculus, we refer to its dual with respect to the Hilbert-Schmidt inner product as the (negative) ***divergence operator***, and this is

$$\nabla_L^* : \mathscr{S}^N \to \mathscr{H}, Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \mapsto \sum_{k}^{N} L_k Y_k - Y_k L_k, \quad (10)$$

i.e., ∇_L^* is defined by means of the identity

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

This notion of gradient operator is motivated by the Lindblad equation in quantum mechanics [21]. The operator ∇_L depends on the choice of the basis L, which is assumed to be fixed throughout. There is no standard way of choosing L. A standing assumption, is that

the null space of ∇_L, denoted by $\ker(\nabla_L)$, contains only scalar multiples of the identity matrix I.

IV. Wassertein-1 distance for density matrices

In this section, we show that both (3) and (4) have natural counterparts for density matrices, i.e. matrices in $\mathscr{D}$. This setup obviously works for matrices in $\mathscr{H}_+$ of equal trace.

We treat (3) as our starting definition and define the W_1 distance in the space of density matrices $\mathscr{D}$ as

$$W_1(\rho_0,\rho_1):=\sup_{f\in\mathscr{H}}\{\mathrm{tr}[f(\rho_1-\rho_0)]|\|\nabla_L f\|\le 1\}. \tag{11}$$

Here $\|\cdot\|$ is the *operator norm*, i.e., the maximum of the singular values. The above is well-defined since by assumption, the null space of ∇_L is spanned by the identity matrix I. This should be compared to the connes spectral distance [30], which is given by

$$d_D(\rho_0,\rho_1)=\sup_{f\in\mathscr{H}}\{\mathrm{tr}[f(\rho_1-\rho_0)]|\|[D,f]\|\le 1\},$$

where $[D, f] = Df - fD$ denotes the Lie bracket. We next show that the dual of (11) is

$$\hat{W}_1(\rho_0,\rho_1)=\inf_{u\in\mathscr{C}^N}\left\{\|u\|_*|\rho_1-\rho_0-\frac{1}{2}\nabla_L^*(u-u_*)=0\right\}, \tag{12}$$

which is the counterpart of (4). Here $\|\cdot\|_*$ denotes the nuclear norm [31], that is, the sum of the singular values. In particular, we have the following theorems.

Theorem 2

Notation as above. Then

$$W_1(\rho_0,\rho_1)=\hat{W}_1(\rho_0,\rho_1).$$

Proof 1

We start from (12) and use the fact that

$$\|u\|_*=\sup_{g\in\mathscr{C}^N,\|g\|\le 1}\Re\langle u,g\rangle=\sup_{\|g\|\le 1}\frac{1}{2}(\langle u,g\rangle+\langle g,u\rangle).$$

It follows

$$\begin{aligned}\hat{W}_1 &= \inf_u \sup_{f\in\mathscr{H}} \left\{ \|u\|_* + \langle f, \rho_1 - \rho_0 - \frac{1}{2}\nabla_L^*(u - u_*)\rangle \right\} \\ &= \inf_u \sup_{f,\|g\|\le 1} \{\Re\langle u, g\rangle - \Re\langle u, \nabla_L f\rangle + \langle f, \rho_1 - \rho_0\rangle\} \\ &= \inf_u \sup_{f,\|g\|\le 1} \{\Re\langle u, g - \nabla_L f\rangle + \langle f, \rho_1 - \rho_0\rangle\} \ge \sup_{f,\|g\|\le 1} \inf_u \{\Re\langle u, g - \nabla_L f\rangle + \langle f, \rho_1 - \rho_0\rangle\} \\ &= \sup_{f,\|g\|\le 1} \{\langle f, \rho_1 - \rho_0\rangle | g = \nabla_L f\} \\ &= \sup_f \{\langle f, \rho_1 - \rho_0\rangle | \|\nabla_L f\| \le 1\} \\ &= W1(\rho_0, \rho_1).\end{aligned}$$

This implies that (11) and (12) are dual to each other. Since both of them are strictly feasible, the duality gap is zero. Therefore $W_1(\rho_0, \rho_1) = \hat{W}_1(\rho_0, \rho_1)$.

Theorem 3

The W_1 distance defined as in (11) is a metric on the space of density matrices $\mathscr{D}$.

Proof 2

Obviously $W_1(\rho_0, \rho_1) \ge 0$ holds with equality if and only if $\rho_0 = \rho_1$. The symmetric property that $W_1(\rho_0, \rho_1) = W_1(\rho_1, \rho_0)$ is also clear from the definition. Here we prove the triangle inequality. That is, for any $\rho_0, \rho_1, \rho_2 \in \mathscr{D}$, we have

$$W_1(\rho_0, \rho_2) \le W_1(\rho_0, \rho_1) + W_1(\rho_1, \rho_2).$$

It is easier to see this from the dual formulation (12). Let u_1, u_2 be the optimal fluxes for (ρ_0, ρ_1) and (ρ_1, ρ_2) respectively. Then $u_1 + u_2$ is a feasible flux for (ρ_0, ρ_2), namely,

$$\rho_2 - \rho_0 - \frac{1}{2}\nabla_L^*(u_1 + u_2 - u_{1*} - u_{2*}) = 0.$$

It follows that

$$W_1(\rho_0, \rho_2) \le \|u_1 + u_2\|_* \le \|u_1\|_* + \|u_2\|_*,$$

which completes the proof.

V. Wasserteln-1 distance: the unbalanced case

In this section, we extend the definition of Wasserstein-1 distance to the space of nonnegative matrices $\mathscr{H}_+$, i.e., we remove the constraint of both matrices having equal traces. Compare also with some very interesting recent work [28] on fast computational methods for $W1$ in the unbalanced scalar case.

In order to compare matrices of unequal trace we relax the constraint in (12), which forces $\mathrm{tr}(\rho_0) = \mathrm{tr}(\rho_1)$, by introducing a "source" term $v \in \mathscr{H}$. That is, we replace our continuity equation (12) with

$$\rho_1 - \rho_0 - \frac{1}{2}\nabla_L^*(u - u_*) - v = 0. \quad (13)$$

With this added source, we define a Wasserstein-1 distance in $\mathscr{H}_+$ as follows. Given $\rho_0, \rho_1 \in \mathscr{H}_+$, we define

$$V_1(\rho_0, \rho_1) = \inf_{\substack{u \in \mathscr{C}^N \\ v \in \mathscr{H}}} \{\|u\|_* + \alpha\|v\|_* | \rho_1 - \rho_0 - \frac{1}{2}\nabla_L^*(u - u_*) - v = 0\}. \quad (14)$$

Here $\alpha > 0$ measures the relative significance between u and v.

Another natural way to compare $\rho_0, \rho_1 \in \mathscr{H}_+$ is by finding $\mu, \nu \in \mathscr{H}_+$ having equal trace that are close to ρ_0, ρ_1 in some norm (here taken to be the nuclear norm), as well as close to one another. More specifically, we seek μ, ν to minimize

$$W_1(\mu, v) + \alpha\|\rho_0 - \mu\|_* + \alpha\|\rho_1 - \nu\|_*. \quad (15)$$

Putting the two terms together we obtain the following definition of Wasserstein-1 distance

$$\hat{V}_1(\rho_0, \rho_1) = \inf_{\substack{u \in \mathscr{C}^N \\ \mu, \nu \in \mathscr{H}_+}} \|u\|_* + \alpha\|\rho_0 - \mu\|_* + \alpha\|\rho_1 - \nu\|_* \quad (16a)$$

$$\nu - \mu - \frac{1}{2}\nabla_L^*(u - u_*) = 0, \quad (16b)$$

$$\mathrm{tr}(\mu) = \mathrm{tr}(\nu). \quad (16c)$$

It turns out these two relaxations of W_1 are in fact equivalent.

Theorem 4

With notation and assumptions as above,

$$V_1(\rho_0, \rho_1) = \hat{V}_1(\rho_0, \rho_1). \quad (17)$$

Proof 3

Clearly, $\hat{V}_1(\rho_0, \rho_1) \geq V_1(\rho_0, \rho_1)$. On the other hand, let u, v be a minimizer of (14), and $v = v_1 - v_0$ with $v_0, v_1 \in \mathscr{H}_+$ i.e., v_0, v_1 are the negative and positive parts of v respectively, then $\mu = \rho_0 + v_0$, $\nu = \rho_1 + v_1$ together with u is a feasible solution to (16). With this,

$$\hat{V}_1(\rho_0, \rho_1) \leq \|u\|_* + \alpha\|\rho_0 - \mu\|_* + \alpha\|\rho_1 - \nu\|_* = \|u\|_* + \alpha\|v_0\|_* + \alpha\|v_1\|_* = \|u\|_* + \alpha\|v\|_*,$$

which implies that $\hat{V}_1(\rho_0, \rho_1) \leq V_1(\rho_0, \rho_1)$. This completes the proof.

Theorem 5

The formula (14) defines a metric on $\mathscr{H}_+$.

Proof 4

The proof follows exactly the same lines as in Theorem 3.

Using the technique of Lagrangian multipliers one can deduce the dual formulation of (12) and establish the following:

Theorem 6

Notation as above. Then

$$V_1(\rho_0, \rho_1) = \sup_{f \in \mathscr{H}} \{\mathrm{tr}[f(\rho_1 - \rho_0)] | \|\nabla_L f\| \leq 1, \|f\| \leq \alpha\}. \tag{18}$$

Proof 5

Straight calculation gives

$$
\begin{aligned}
&V_1(\rho_0,\rho_1)\\
&\quad=\inf_{u,v}\sup_{f\in\mathscr{H}}\{\|u\|_*\\
&\quad+\alpha\|v\|_*+\langle f,\rho_1\\
&\qquad-\rho_0-\frac{1}{2}\nabla_L^*(u-u_*)-u\rangle\}\\
&\quad=\inf_{u,v}\sup_{f,\|g\|\le 1,\|h\|\le 1}\{\Re\langle u,g\rangle\\
&\quad+\alpha\Re\langle v,h\rangle\\
&\quad+\langle f,\rho_1\\
&\quad-\rho_0-\frac{1}{2}\nabla_L^*(u\\
&\qquad-u_*)-v\rangle\}\ge\sup_{f,\|g\|\le 1,\|h\|\le 1}\inf_{u,v}\{\Re\langle u,g\\
&\quad-\nabla_L f\rangle\\
&\quad+\Re\langle v,\alpha h\\
&\quad-f\rangle+\langle f,\rho_1\\
&\qquad-\rho_0\rangle\}\\
&\quad=\sup_f\{\langle f,\rho_{1,}\\
&\qquad-\rho_0\rangle|\|\nabla_L f\|\le 1,\|f\|\le\alpha\}.
\end{aligned}
$$

This together with the strong duality completes the proof.

VI. Wasserstein-1 distance for matrix-valued densities

With little effort we are able to generalize the definition of Wasserstein-1 distance to the space of matrix-valued densities. Examples of matrix-valued densities include power spectra of multivariate time series, stress tensors, diffusion tensors and so on, and hence our motivation in considering matrix-valued distribution on possibly more than a one dimensional spatial coordinates.

Given two matrix-valued densities ρ_0, ρ_1 satisfying

$$\int_{\mathbb{R}^m}\mathrm{tr}(\rho_0(x))dx=\int_{\mathbb{R}^m}\mathrm{tr}(\rho_1(x))dx, \quad (19)$$

we can define their Wasserstein-1 distance as

$$W_1(\rho_0,\rho_1):=\sup_{f\in\mathscr{H}}\left\{\int_{\mathbb{R}^m}\mathrm{tr}[f(x)(\rho_1(x)-\rho_0(x))]dx\,\middle|\,\left\|\begin{bmatrix}\nabla_x f\\ \nabla_L f\end{bmatrix}\right\|\le 1\right\},$$

or through its dual

$$W_1(\rho_0,\rho_1)=\inf_{\substack{u_1\in\mathscr{C}^m\\ u_2\in\mathscr{C}^N}}\{\int_{\mathbb{R}^m}\left\|\begin{bmatrix}u_1(x)\\ u_2(x)\end{bmatrix}\right\|_* dx|\rho_1-\rho_0+\frac{1}{2}\nabla_x\cdot(u_1+u_{1*})-\frac{1}{2}\nabla_L^*(u_2-u_{2*})=0\}.$$

For more general densities where condition (19) may not be valid, we define

$$V_1(\rho_0,\rho_1):=\sup_{f\in\mathscr{H}}\left\{\int_{\mathbb{R}^m}\mathrm{tr}[f(x)(\rho_1(x)-\rho_0(x))]dx\right|\left\|\begin{bmatrix}\nabla_x f\\ \nabla_L f\end{bmatrix}\right\|\le 1,\|f\|\le\alpha\Big\},$$

or, equivalently,

$$V_1(\rho_0,\rho_1)=\inf_{\substack{u_1\in\mathscr{C}^m\\ u_2\in\mathscr{C}^N,v\in\mathscr{H}}}\left\{\int_{\mathbb{R}^m}\|\begin{bmatrix}u_1(x)\\ u_2(x)\end{bmatrix}\|_*+\alpha\|v\|_* dx|\rho_1-\rho_0+\frac{1}{2}\nabla_x\cdot(u_1+u_{1*})\frac{1}{2}\nabla_x^*(u_2-u_{2*})-v=0\right\}.$$

One can introduce positive coefficients $\beta_1 > 0$, $\beta_2 > 0$ to trade-off the relative importance of u_1 and u_2 in establishing correspondence between the two distributions as follows:

$$V_1(\rho_0,\rho_1):=\sup_{f\in\mathscr{H}}\left\{\int_{\mathbb{R}^m}\mathrm{tr}[f(x)(\rho_1(x)-\rho_0(x))]dx\right|\left\|\begin{bmatrix}\beta_1\nabla_x f\\ \beta_2\nabla_L f\end{bmatrix}\right\|\le 1,\|f\|\le\alpha\Big\},$$

or, equivalently,

$$V_1(\rho_0,\rho_1)=\inf_{\substack{u_1\in\mathscr{C}^m\\ u_2\in\mathscr{C}^N,v\in\mathscr{H}}}\left\{\int_{\mathbb{R}^m}\left\|\begin{bmatrix}u_1(x)\\ u_2(x)\end{bmatrix}\right\|_*+\alpha\|v\|_* dx|\rho_1-\rho_0+\frac{\beta_1}{2}\nabla_x\cdot(u_1+u_{1*})-\frac{\beta_2}{2}\nabla_L^*(u_2-u_{2*})-v=0\right\}.$$

Remark 7

By splitting the penalty on the two fluxes u_1, u_2 into two terms $\|u_1\|_*+\|u_2\|_*$, we attain an alternative definition of Wasserstein-1 for matrix-valued densities, as

$$W_1(\rho_0,\rho_1)=\inf_{\substack{u_1\in\mathscr{C}^m\\ u_2\in\mathscr{C}^N}}\left\{\int_{\mathbb{R}^m}\|u_1(x)\|_*+\|u_2(x)\|_* dx|\rho_1-\rho_0+\frac{1}{2}\nabla_x\cdot(u_1+u_{1*})-\frac{1}{2}\nabla_L^*(u_2-u_{2*})=0\right\}$$

with dual formulation

$$W_1(\rho_0,\rho_1)=\sup_{f\in\mathscr{H}}\{\int_{\mathbb{R}^m}\mathrm{tr}[f(x)(\rho_1(x)-\rho_0(x))]dx|\|\nabla_x f\|\le 1,\|\nabla_L f\|\le 1\}.$$

VII. Example

We use our framework to compare power spectra of multivariate time series (in discrete time). Evidently, the distance between two power spectra induces a distance between

corresponding linear modeling filters and, thereby, can be used to compare (stable) MIMO systems [18].

Consider the three power spectra as shown in Figure 1 (in different colors). What is shown in the three subplots are power spectra of two time series (in subplots (a) and (c)) and their cross-spectrum (in subplot (b)) as functions of time (the phase of the cross spectra are not shown). Thus, the three different colors represent the three different matrixvalued power spectra given by:

$$\rho_0(\theta)=\begin{bmatrix} 1 & 0.4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0.01 & 0 \\ 0 & \frac{0.7}{|a_0(e^{j\theta})|^2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0.4 & 1 \end{bmatrix}$$

$$\rho_1(\theta)=\begin{bmatrix} 1 & 0.5 \\ 0.5e^{j\theta} & 1 \end{bmatrix}\begin{bmatrix} \frac{0.5}{|a_1(e^{j\theta})|^2} & 0 \\ 0 & \frac{0.5}{|a_1(e^{j\theta})|^2} \end{bmatrix}\begin{bmatrix} 1 & 0.5e^{-j\theta} \\ 0.5 & 1 \end{bmatrix}$$

$$\rho_2(\theta)=\begin{bmatrix} 1 & 0 \\ 0.4e^{j\theta} & 1 \end{bmatrix}\begin{bmatrix} \frac{0.2}{|a_2(e^{j\theta})|^2} & 0 \\ 0 & 0.02 \end{bmatrix}\begin{bmatrix} 1 & 0.4e^{-j\theta} \\ 0 & 1 \end{bmatrix}$$

where

$$a_0(z)=(1-1.9\cos(\frac{\pi}{6})z-0.95^2z^2)\times(1-1.5\cos(\frac{\pi}{3})z+0.75^2z^2)$$

$$a_1(z)=(1-1.9\cos(\frac{2\pi}{3})z-0.95^2z^2)\times(1-1.5\cos(\frac{5\pi}{8})z+0.75^2z^2)$$

$$a_2(z)=(1-1.9\cos(\frac{5\pi}{12})z-0.95^2z^2)\times(1-1.5\cos(\frac{\pi}{2})z+0.75^2z^2).$$

The distances between the each pair for different β_1, β_2 values, $\alpha = 1$, for the particular choice $L = [L_1, L_2]$ with

$$L_1=\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, L_2=\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},$$

are tabulated in Table I. We observe that when the penalty on the rotation part is large ($\beta_1 \gg \beta_2$), we have $V_1(\rho_0, \rho_2) > V_1(\rho_0, \rho_1)$ and $V_1(\rho_0, \rho_2) > V_1(\rho_2, \rho_1)$. On the other hand, when the penalty on translation is large relative to the cost of rotation ($\beta_1 \ll \beta_2$), we have $V_1(\rho_0, \rho_1) > V_1(\rho_0, \rho_2)$ and $V_1(\rho_0, \rho_1) > V_1(\rho_1, \rho_2)$. These findings are in agreement with the

intuition when observing the relative frequency directionality of power in the three spectra. More specifically, ρ_1 requires a significant drift in directionality before we can match it with the other two, while this is less important when comparing ρ_0 and ρ_2. For this latter case, it is the actual frequency where the power resides that distinguishes the two while the directionality is more in agreement.

What this example underscores is the ability of the metric to be tailored to applications where we need to trade off and compromise, in a principled way, between two vastly different features of matrix-valued distributions, i.e., spatial location versus directionality of the "intensity." What was achieved in this paper is the construction of a suitable and easily computable metric that can be utilized for this purpose.

VIII. Future research

We introduced generalization of the scalar W_1 distance to matrices and matrix-valued measures. This new metric, W_1, is computationally simpler and more attractive than earlier metrics, based on quadratic cost criteria. In fact, our "dual of the dual" formulation makes the metric especially attractive when comparing matrix-valued data on a discrete space (graph, network). In particular, the primal-dual algorithm developed in [28], [29] for scalar W_1 problems based on the "dual of the dual" formulation has proved to be efficient and reliable. Our next step will be generalizing it to matrix-valued densities.

We note that the Wasserstein 1-metric has been used as a tool in defining *curvature* [32] and in analyzing the *robustness* of complex networks derived from scalar-valued data [24], [25]. The formalism presented in the current work, suggests alternative notions of curvature and robustness when the nodes of a network carry matrix-valued data, e.g., in diffusion tensor imaging. We plan to pursue such issues in future work.

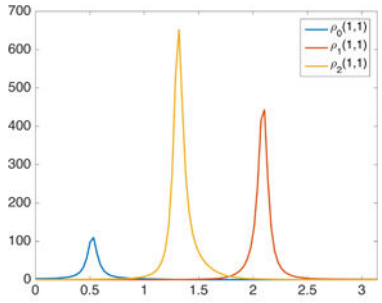
Acknowledgments

This project was supported by AFOSR grants (FA9550-15-1-0045 and FA9550-17-1-0435), 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), and National Institutes of Health (P30-CA-008748 and 1U24CA18092401A1).

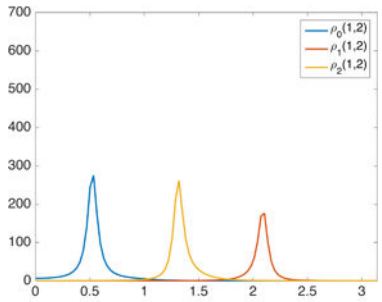
References

1. Rachev ST, Rüschendorf L. Mass Transportation Problems: Volume I: Theory Springer. 1998; 1
2. Benamou JD, Brenier Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numerische Mathematik. 2000; 84(3):375–393.
3. Villani C. Topics in Optimal Transportation American Mathematical Soc. 2003; 58
4. Tannenbaum E, Georgiou T, Tannenbaum A. Signals and control aspects of optimal mass transport and the boltzmann entropy. Proceedings of the 49th IEEE Conference on Decision and Control IEEE. 2010:1885–1890.
5. Mueller M, Karasev P, Kolesov I, Tannenbaum A. Optical flow estimation for flame detection in videos. IEEE Trans on Image Processing. 2013; 22(7):2786–2797.
6. Pavon M, Wakolbinger A. On free energy, stochastic control, and Schrödinger processes. in Modeling, Estimation and Control of Systems with Uncertainty Springer. 1991:334–348.
7. Mikami T, Thieullen M. Optimal transportation problem by stochastic optimal control. SIAM Journal on Control and Optimization. 2008; 47(3):1127–1139.

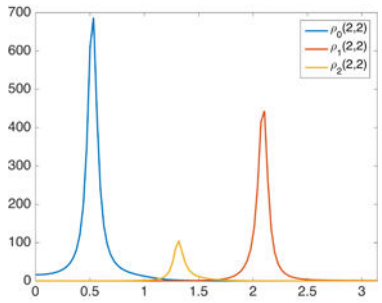
8. Leonard C. A survey of the Schrödinger problem and some of its connections with optimal transport. Discrete Contin Dyn Syst A. 2014; 34(4):1533–1574.
9. Dai Pra P. A stochastic control approach to reciprocal diffusion processes. Applied Mathematics and Optimization. 1991; 23(1):313–329.
10. Chen Y, Georgiou TT, Pavon M. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. Journal of Optimization Theory and Applications. 2016; 169(2):671–691.
11. Chen Y, Georgiou T, Pavon M. Optimal transport over a linear dynamical system. arXiv preprint arXiv:1502.01265, IEEE Trans on Automatic Control, to appear. 2017
12. Chen Y, Georgiou TT, Pavon M. Optimal steering of a linear stochastic system to a final probability distribution, Part I. IEEE Trans on Automatic Control. 2016; 61(5):1158–1169.
13. Chen Y, Georgiou TT, Pavon M. Optimal steering of a linear stochastic system to a final probability distribution, Part II. IEEE Trans on Automatic Control. 2016; 61(5):1170–1180.
14. Chen Y, Georgiou T, Pavon M, Tannenbaum A. Robust transport over networks. IEEE Trans on Automatic Control. 2016
15. McCann RJ. A convexity principle for interacting gases. Advances in mathematics. 1997; 128(1): 153–179.
16. Jordan R, Kinderlehrer D, Otto F. The variational formulation of the Fokker-Planck equation. SIAM Journal on Mathematical Analysis. 1998; 29(1):1–17.
17. Otto F. The geometry of dissipative evolution equations: the porous medium equation. Communications in Partial Differential Equations. 2001
18. Ning L, Georgiou TT, Tannenbaum A. On matrix-valued Monge-Kantorovich optimal mass transport. IEEE Trans Automatic Control. 2015; 60(2):373–382. [PubMed: 26997667]
19. Ning L, Georgiou TT. Metrics for matrix-valued measures via test functions. Proceedings of the 53rd IEEE Conference on Decision and Control IEEE. 2014:2642–2647.
20. Carlen EA, Maas J. Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. arXiv:1609.01254. 2016
21. Chen Y, Georgiou TT, Tannenbaum A. Matrix optimal mass transport: a quantum mechanical approach. arXiv:1610.03041. 2016
22. Mittnenzweig M, Mielke A. An entropic gradient structure for lindblad equations and generic for quantum systems coupled to macroscopic models. arXiv:1609.05765. 2016
23. Evans, LC., Gangbo, W. Differential equations methods for the Monge-Kantorovich mass transfer problem. Vol. 653. American Mathematical Soc; 1999.
24. Sandhu R, Georgiou T, Reznik E, Zhu L, Kolesov I, Senbabaoglu Y, Tannenbaum A. Graph curvature for differentiating cancer networks. Scientific reports. 2015; 5:12323. [PubMed: 26169480]
25. Sandhu RS, Georgiou TT, Tannenbaum AR. Ricci curvature: An economic indicator for market fragility and systemic risk. Science advances. 2016; 2(5):e1501495. [PubMed: 27386522]
26. Wang C, Jonckheere E, Banirazi R. Wireless network capacity versus Ollivier-Ricci curvature under heat-diffusion (hd) protocol. Proceedings of the 2014 American Control Conference IEEE. 2014:3536–3541.
27. Chen Y, Cruz FD, Sandhu R, Kung A, Mundi P, Deasy J, Tannenbaum A. Pediatric sarcoma data forms a unique cluster measured via the earth mover's distance. bioRxiv. 2017:116384.
28. Li W, Yin P, Osher S. A fast algorithm for unbalanced L1 Monge-Kantorovich problem. CAM report. 2016
29. Li W, Ryu EK, Osher S, Yin W, Gangbo W. A parallel method for earth mover's distance. 2017
30. Connes, A. Noncommutative geometry. Academic Press; 1994.
31. Candès EJ, Tao T. The power of convex relaxation: Near-optimal matrix completion. IEEE Transactions on Information Theory. 2010; 56(5):2053–2080.
32. Ollivier Y. Ricci curvature of Markov chains on metric spaces. Journal of Functional Analysis. 2009; 256(3):810–864.



(a) $\rho(1,1)$



(b) $\rho(1,2)$



(c) $\rho(2,2)$

Fig. 1.
Power spectra

TABLE I

Distances between power spectra

	ρ_0, ρ_1	ρ_1, ρ_2	ρ_0, ρ_2
$\beta_1 = 10, \beta_2 = 1$	77.85	77.76	137.36
$\beta_1 = 1, \beta_2 = 1$	249.40	162.03	199.78
$\beta_1 = 1, \beta_2 = 10$	210.93	110.25	113.46