SIAM J. APPL. MATH.
Vol. 78, No. 3, pp. 1682–1696

VECTOR-VALUED OPTIMAL MASS TRANSPORT*

YONGXIN CHEN†, TRYPHON T. GEORGIOU‡, AND ALLEN TANNENBAUM§

Abstract. We introduce the problem of transporting vector-valued distributions. In this, a salient feature is that mass may flow between vectorial entries as well as across space (discrete or continuous). The theory relies on a first step taken to define an appropriate notion of optimal transport on a graph. The corresponding distance between distributions is readily computable via convex optimization and provides a suitable generalization of Wasserstein-type metrics. Building on this, we define Wasserstein-type metrics on vector-valued distributions supported on continuous spaces as well as graphs. Motivation for developing vector-valued mass transport is provided by applications such as color image processing, multimodality imaging, polarimetric radar, as well as network problems where resources may be vectorial.

Key words. optimal mass transport, geometry of distributions, image processing, signal processing

AMS subject classifications. 49J20, 49J15, 90B06, 94A08, 49M30

DOI. 10.1137/17M1130897

1. Introduction.

The Monge–Kantorovich optimal mass transport (OMT) theory has witnessed a fast pace of new developments; see [23, 28] for extensive lists of references. These contributions were driven by a multitude of applications in physics, geosciences, economics, and probability. Some of the notable advances include the concept of displacement interpolation [20], links to the geometry of spaces [2, 19, 22, 29], and a fluid dynamic reformulation [3]. In our own work, image analysis and spectral analysis of time series provided starting points (e.g., [13, 14]) and, more recently, problems in stochastic control, quantum information, and matrix-valued distributions [5, 15, 21, 27] provided additional research directions. The present paper continues the work of [4, 6] by proposing a transportation problem for vector-valued distributions.

A salient feature of vector-valued distributions is the possibility of the transfer of "mass" from one vectorial entry to another. Physical examples include color image scenes where the vectorial distribution captures color intensities, which may continuously shift with lighting conditions. Alternatively, polarimetric data provide an analogous example where mass represents power at different polarizations detected at the locations of a sensor array. As another example, the flow of mass between vector entries may represent mutation of coexisting population species.

The proposed framework may have far-reaching consequences in, for example, combining genomic and proteomic networks, and, in general, fusion of vectorial data

*Received by the editors May 18, 2017; accepted for publication (in revised form) April 20, 2018; published electronically June 14, 2018.
http://www.siam.org/journals/siap/78-3/M113089.html
Funding: This work was supported by AFOSR grants (FA9550-15-1-0045 and FA9550-17-1-0435), an ARO grant (W911NF-17-1-0429), an NSF grant (grant ECCS-1509387), 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 grants from the National Institutes of Health (P30-CA-008748, 1U24CA18092401A1, R01-AG048769).
†Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 (Yongchen@iastate.edu).
‡Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697 (tryphon@uci.edu).
§Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY 11794 (allen.tannenbaum@stonybrook.edu).

supported on a graph. While in some of these examples the total mass may not be preserved, in the present work we will restrict our attention to the case where it is. Thus, we seek a suitable continuity equation that allows trading off mass between vectorial entries of a distribution, on a continuous or a discrete space (graph), and we develop a geometric framework that would allow constructing geodesic flows between snapshots of such distributions.

In order to formulate transport between vectorial entries, we begin with a new notion of transport on weighed undirected graphs in the spirit of Erbar and Maas [11]. A starting point in [11] is to devise a suitable continuity equation for probability measures on the nodes of a weighted graph (Markov chain). The formulation in our paper differs from that in [11], and the corresponding transport problem has the advantage of being reducible to one of convex optimization. Both [11] and our formulation were inspired by the Benamou–Brenier theory [3], where the OMT with quadratic cost is recast as the problem to minimize flow kinetic energy (i.e., an action integral). The present work builds on [4, 6], extending the Wasserstein theory to densities and mass distributions on more general spaces. Having as a first step a Benamou–Brenier theory on graphs, the methodology allows us to define a notion of vector-valued transport and corresponding distance between vector-valued densities on discrete or continuous spaces. As with the (weighted) graph case, the transport distance that we define on vector-valued densities may be reduced to a convex optimization problem.

We now outline the remainder of this article. In section 2, we sketch needed background from the classical theory of optimal mass transport that motivates our generalization. In section 3, we describe the proposed Wasserstein-2 metric on an undirected weighted graph. Further, we remark on the Wasserstein-1 type of metric on a weighted graph. In sections 4 and 5, we formulate the new Wasserstein distance on vector-valued densities that are supported, first on the Euclidean spaces and then on graphs. In section 6, we give several examples illustrating the idea of a vector-valued optimal mass transport, and finally we conclude in section 7 with an outline of possible applications of the theory and future research directions.

2. Preliminaries on optimal mass transport. The mass transport problem was first formulated by Gaspar 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. This problem was given a modern formulation in the work of Kantorovich in the form of a linear program, and it is now known as the *Monge–Kantorovich problem.* See [12, 16, 23, 28] for all details as well as extensive lists of references.

Herein, we focus mainly the case where the transportation cost is quadratic in the distance. The respective optimization problem

$$W_p(\rho_0,\rho_1)^p := \inf_{\pi\in\Pi(\rho_0,\rho_1)} \int_{\mathbb{R}^N\times\mathbb{R}^N} \|x-y\|^p \pi(dx,dy) \tag{1}$$

for $p = 2$, where $\|\cdot\|$ denotes Euclidean distance and $\Pi(\rho_0,\rho_1)$ represents the set of all couplings between two ρ_0 and ρ_1 nonnegative probability density functions on $\mathbb{R}^N$ (i.e., the set of joint probability distributions having ρ_0 and ρ_1 as respective marginals), defines the so-called Wasserstein-2 distance between the two densities, or more generally, between measures.

In this case, where the cost is quadratic (i.e., $p = 2$), the transport problem admits a dynamic reformulation [3] that is especially powerful, and the space of densities $\mathcal{D} := \{\rho \geq 0 \mid \int \rho\, dx = 1\}$ admits, essentially, a Riemannian structure [22]. The Benamou–Brenier reformulation identifies the Wasserstein-2 distance with the integral of the

kinetic energy (action integral) along a geodesic flow that links the two marginals, namely,

$$W_2(\rho_0,\rho_1)^2 = \inf_{\rho,v} \int \int_0^1 \rho(t,x)\|v(t,x)\|^2\, dt\, dx \tag{2a}$$

over all time varying densities ρ and vector fields v satisfying the continuity equation and boundary conditions

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

$$\rho(0,\cdot) = \rho_0,\ \rho(1,\cdot) = \rho_1. \tag{2c}$$

Interestingly, when expressed in terms of density ρ and flux $m = \rho v$, the minimization problem in (2a) becomes convex while (2b) and (2c) turn into linear constraints. For the optimal pair (ρ, v), the vector field turns out to be the gradient $v = \nabla_x g$ of a function g, hence it is "rot-free." Vector fields v of this form can be identified with tangent directions of $\mathcal{D}$, i.e., elements of the tangent space

$$T_\rho \mathcal{D} \cong \left\{ \delta : \int \delta = 0 \right\},$$

as follows. Under suitable assumptions on differentiability for $\rho \in \mathcal{D}$ and $\delta \in T_\rho \mathcal{D}$, we solve the Poisson equation

$$\delta = -\nabla_x \cdot (\rho \nabla_x g) \tag{3}$$

to obtain a convex function g_δ and thereby the vector field ∇g_δ. In this way the space $\mathcal{D}$ can be endowed with a Riemannian structure (see [22, 28]) via

$$\langle \delta_1, \delta_2 \rangle_\rho := \int \rho \nabla_x g_{\delta_1} \cdot \nabla_x g_{\delta_2} dx, \tag{4}$$

which has the aforementioned kinetic energy interpretation. This inner product induces precisely the Wasserstein distance as the geodesic distance between the two marginals in (2c).

Remark 1 (Wasserstein-1 metric). Interestingly, the case of linear cost (i.e., $p = 1$), can also be cast as the flux minimization problem

$$W_1(\rho_0,\rho_1) = \inf_m \int \|m\|\, dx,$$
$$\rho_1 - \rho_0 + \nabla_x \cdot m = 0,$$

as it can be shown that the flux $m = \rho v$ remains invariant with time [12]. For this case, there is also an alternative expression through a dual formulation,

$$W_1(\rho_0,\rho_1) = \sup_f \left\{ \int f(\rho_1 - \rho_0) dx \mid \|\nabla_x f\| \le 1 \right\},$$

in terms of test functions; see [12, 23, 28, 18].

3. Wasserstein metric on weighted graphs. Following the Benamou–Brenier viewpoint on Wasserstein distances, our first task is to develop an analogous notion of transportation distances on graphs. To this end, we consider a connected, positively weighted, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ with n nodes labeled as i, with $1 \le i \le n$, and m edges. We consider the set of probability masses on $\mathcal{G}$ that we will denote by $\mathcal{D}$; an element $\rho \in \mathcal{D}$ may be regarded as a column vector $(\rho_1, \dots, \rho_n)^T$, with $\rho_i \ge 0$ for $1 \le i \le n$ and

$$\sum_{i=1}^{n} \rho_i = 1.$$

We denote the (open) interior of $\mathcal{D}$ by $\mathcal{D}_+$.

The standard heat equation on $\mathcal{G}$,

$$\dot{\rho} = L\rho = -DWD^T\rho,$$

where $L, D, W = \operatorname{diag}\{w_1, \dots, w_m\}$ are the graph-Laplacian, incidence, and weight matrices, respectively, can also be written in the more familiar (from calculus)

$$\dot{\rho} = \Delta_{\mathcal{G}}\rho \tag{5}$$

by defining

$$\Delta_{\mathcal{G}} := -\nabla_{\mathcal{G}}^*\nabla_{\mathcal{G}},$$

where

$$\nabla_{\mathcal{G}} : \mathbb{R}^n \to \mathbb{R}^m,\ x \mapsto W^{1/2}D^T x$$

denotes the gradient operator and

$$\nabla_{\mathcal{G}}^* : \mathbb{R}^m \to \mathbb{R}^n,\ y \mapsto DW^{1/2}y$$

denotes its dual. More generally, if we let the entries of $u(t) \in \mathbb{R}^m$ represent flux along respective edges, we can express the continuity equation in the form

$$\dot{\rho} - \nabla_{\mathcal{G}}^* u = 0. \tag{6}$$

Evidently, the flux $u = -\nabla_{\mathcal{G}}\rho$ gives (5). Also note that since the row vector consisting of all 1's lies in the left kernel of the incidence matrix, mass is preserved by (6).

To carry out our program, we need to express the flux u in the form of a momentum "ρv" as in [3]. However, the flux is supported on the edges $\mathcal{E}$ of the graph whereas the mass is supported on the set of nodes $\mathcal{V}$, the two sets having different dimensions. In order to overcome this difficulty in a natural manner, we choose to associate the flux along an edge with the mass at the source in the two endpoints. More specifically, the flux along an edge $e_k = (i, j)$, with source i and sink j, consists of two parts. A part that flows out of node i, and another that flows in opposite direction out of node j. Thus, we define a flux $u_k = \rho_i v_k$ out of i and another, $\bar{u}_k = \rho_j \bar{v}_k$ out of j, and represent the total flux as the superposition $\rho_j v_k - \rho_i \bar{v}_k$, while restricting the rates $v_k, \bar{v}_k$ to be nonnegative. Thus, our *continuity equation* for rates $v, \bar{v} \in \mathbb{R}^m_+$ becomes

$$\dot{\rho} - \nabla_{\mathcal{G}}^*((D_2^T\rho) \circ v - (D_1^T\rho) \circ \bar{v}) = 0, \tag{7}$$

where $\circ$ denotes entry-wise multiplication of two vectors. The matrix D_1 is the portion of the incidence matrix D containing 1's (sources), and $D_2 = D_1 - D$ (sinks). In other words, $D_1^T\rho$ is the mass at the source of an edge, and $D_2^T\rho$ is the mass at the sink of

an edge. *The dependence of the flux in* (7) *on* ρ *ensures that the entries of* ρ *remain positive while the fact that the kernel of* D *contains the vector with all ones ensures that the total mass is preserved as well.*

For notational convenience we use $\mu \in \mathcal{D}_+$ and $\nu \in \mathcal{D}_+$ (instead of ρ_0 and ρ_1) to denote the starting and ending mass on nodes. We now define the transport distance between μ and ν as follows:

$$
\begin{aligned}
(8) \qquad W_{2,a}(\mu,\nu)^2 := \inf_{\rho,v,\bar{v}} & \int_0^1 \left\{ v^T((D_2^T\rho)\circ v) + \bar{v}^T((D_1^T\rho)\circ \bar{v}) \right\} dt \\
& \dot{\rho} - \nabla_{\mathcal{G}}^*((D_2^T\rho)\circ v - (D_1^T\rho)\circ \bar{v}) = 0, \\
& v \geq 0, \quad \bar{v} \geq 0, \\
& \rho(0) = \mu, \quad \rho(1) = \nu.
\end{aligned}
$$

It is easy to see that at each time instant, for each k, at most one of the v_k and $\bar{v}_k$ is nonzero. In a similar manner as in the Benamou–Brenier program, (8) can be recast in the form of a convex optimization problem in (momentum) variables $u = (D_2^T\rho)\circ v$, $\bar{u} = (D_1^T\rho)\circ \bar{v}$,

$$
\begin{aligned}
(9) \qquad W_{2,a}(\mu,\nu)^2 = \inf_{\rho,u,\bar{u}} & \int_0^1 \left\{ u^T \operatorname{diag}(D_2^T\rho)^{-1} u + \bar{u}^T \operatorname{diag}(D_1^T\rho)^{-1}\bar{u} \right\} dt \\
& \dot{\rho} - \nabla_{\mathcal{G}}^*(u - \bar{u}) = 0, \\
& u \geq 0, \quad \bar{u} \geq 0, \\
& \rho(0) = \mu, \quad \rho(1) = \nu.
\end{aligned}
$$

It is straightforward to see that the right-hand side in (9) is in general positive and vanishes only when $\mu = \nu$. It is also straightforward to see that $W_{2,a}$ satisfies the triangle inequality. However, in general, $W_{2,a}(\mu,\nu) \neq W_{2,a}(\nu,\mu)$, therefore $W_{2,a}$ is only a quasimetric. Yet, it endows $\mathcal{D}_+$ with a Finsler metric type structure,

$$
\begin{aligned}
(10) \qquad W_{2,a}(\mu,\mu+\delta)^2 = \inf_{\rho,v,\bar{v}} & v^T((D_2^T\mu)\circ v) + \bar{v}^T((D_1^T\mu)\circ \bar{v} \\
& \delta - \nabla_{\mathcal{G}}^*((D_2^T\mu)\circ v - (D_1^T\mu)\circ \bar{v}) = 0, \\
& v \geq 0, \quad \bar{v} \geq 0,
\end{aligned}
$$

for small perturbation δ, and in this, $(\mathcal{D}_+, W_{2,a})$ becomes a length space. In fact, $W_{2,a}$ has a very nice "geodesic" property. Indeed, if $\rho(\cdot)$ is the mass distribution as a function of time obtained by solving (9), then

$$
(11) \qquad W_{2,a}(\rho(s),\rho(t)) = (t-s)W_{2,a}(\mu,\nu)
$$

for any $0 \leq s < t \leq 1$. Finally, $W_{2,a}$ can be extended to $\mathcal{D}$, the closure of $\mathcal{D}_+$, by continuity.

Naturally, one can symmetrize $W_{2,a}$ in the obvious way, by adopting as our metric

$$
\max\{W_{2,a}(\mu,\nu), W_{2,a}(\nu,\mu)\},
$$

which can then be computed by solving two convex optimization problems. (A similar remark holds for the metrics $W_{2,b}$ and $W_{2,c}$ defined later on.)

An alternative way to symmetrize $W_{2,a}$ is to replace the cost function (9) with

$$
\int_0^1 \frac{1}{2} \left\{ u^T(\operatorname{diag}(D_2^T\rho)^{-1} + \operatorname{diag}(D_1^T\rho)^{-1})u + \bar{u}^T(\operatorname{diag}(D_2^T\rho)^{-1} + \operatorname{diag}(D_1^T\rho)^{-1})\bar{u} \right\} dt.
$$

Since the cost terms for u and $\bar{u}$ are symmetric, we can combine the two and drop the nonnegativity requirement $u \geq 0, \bar{u} \geq 0$, to obtain

$$\hat{W}_{2,a}(\mu,\nu)^2 := \inf_{\rho\in\mathcal{D}_+,u} \int_0^1 \left\{u^T(\operatorname{diag}(D_2^T\rho)^{-1} + \operatorname{diag}(D_1^T\rho)^{-1})u\right\} dt \tag{12}$$
$$\dot{\rho} - \nabla_{\mathcal{G}}^* u = 0,$$
$$\rho(0) = \mu, \quad \rho(1) = \nu.$$

Positive entries of u represent flow from sources to sinks, while negative entries flow from sinks to sources. This (symmetric) metric induces a Riemannian type structure on $\mathcal{D}_+$, akin to that of standard optimal transport theory on Euclidean spaces [28].

Remark 2. We point out that similar notions was recently considered in [11, 8, 26, 9]. We arrived at the above formulation independently and from a different starting point.

Remark 3 (Gradient flow of entropy). The gradient flow of the entropy functional on probability mass distribution on a graph $\mathcal{G}$ with respect to $\hat{W}_{2,a}$ is given by

$$\dot{\rho} = -\nabla_{\mathcal{G}}^*(A(\rho)^{-1}\nabla_{\mathcal{G}}\log\rho), \tag{13}$$

where $A(\rho) := \operatorname{diag}(D_2^T\rho)^{-1} + \operatorname{diag}(D_1^T\rho)^{-1}$. It represents a nonlinear heat-like equation, to be contrasted with the linear heat equation derived in [11]. To see (13), compute the derivative of the entropy functional $S(\rho) = -\sum_{i=1}^n \rho_i \log\rho_i$ along a curve $\rho(t), t \in [0,1]$ in $\mathcal{D}_+$,

$$\begin{aligned}
-\dot{S}(\rho) &= \sum_{i=1}^n \dot{\rho}_i \log\rho_i + \sum_{i=1}^n \dot{\rho}_i \\
&= \langle \dot{\rho}, \log\rho\rangle \quad (\text{since for each } t,\ \textstyle\sum_{i=1}^n \rho_i(t) = 1) \\
&= \langle \nabla_{\mathcal{G}}^* u, \log\rho\rangle \quad (\text{since } \dot{\rho} = \nabla_{\mathcal{G}}^* u) \\
&= \langle u, \nabla_{\mathcal{G}}\log\rho\rangle,
\end{aligned}$$

and observe that the direction of steepest ascent is along $u = -A(\rho)^{-1}\nabla_{\mathcal{G}}\log\rho$, from which (13) follows.

Remark 4 (Wasserstein-1 distance on graphs). Following up on Remark 1, we sketch a Benamou–Brenier type reformulation of W_1-distances on graphs. Assuming that the entries of $c = (c_1, \ldots, c_m)^T$ are edge weights representing the cost of moving a unit mass across, a W_1 distance between mass distributions μ, ν can be defined as the solution of the min-cost flow problem

$$W_1(\mu,\nu) = \min_u c^T|u| \tag{14}$$
$$\nu - \mu - Du = 0.$$

Alternatively, in terms of "test vectors" f, the expression

$$W_1(\mu,\nu) = \max_f \left\{f^T(\nu-\mu) \ \mid \ \|\nabla_{\mathcal{G}} f\|_\infty \leq 1\right\} \tag{15}$$

has the dual

$$W_1(\mu,\nu) = \min_{\hat{u}} \left\{\|\hat{u}\|_1 \ \mid \ \nu - \mu - \nabla_{\mathcal{G}}^*\hat{u} = 0\right\}. \tag{16}$$

This coincides with (14) by taking $c_i = 1/\sqrt{w_i}$, $1 \le i \le m$, and $u = W^{1/2}\hat{u}$. Finally, we point out that the above has an action minimization formulation following [17]:

$$\begin{aligned} W_1(\mu,\nu) = \inf_{\rho,v,\bar{v}} & \int_0^1 \left\{ c^T((D_2^T\rho)\circ v) + c^T((D_1^T\rho)\circ \bar{v}) \right\} dt \\ & \dot{\rho} - D((D_2^T\rho)\circ v - (D_1^T\rho)\circ \bar{v}) = 0, \\ & v \ge 0, \quad \bar{v} \ge 0, \\ & \rho(0) = \mu, \quad \rho(1) = \nu, \end{aligned}$$

which assumes the convex recast

$$\begin{aligned} W_1(\mu,\nu) = \inf_{\rho,u,\bar{u}} & \int_0^1 \left\{ c^T u + c^T \bar{u} \right\} dt \\ & \dot{\rho} - D(u - \bar{u}) = 0, \\ & u \ge 0, \quad \bar{u} \ge 0, \\ & \rho(0) = \mu, \quad \rho(1) = \nu. \end{aligned}$$

4. Vector-valued densities and transport. We now turn to the main theme of our paper: the introduction of a Wasserstein type metric between vector-valued densities. A vector-valued density $\rho = [\rho_1, \rho_2, \ldots, \rho_M]^T$ on $\mathbb{R}^N$, or on a discrete space, may represent a physical entity that can mutate or be transported between alternative manifestations, e.g., power reflected off a surface at different frequencies or polarizations. While the total power may be invariant (under some lighting conditions), the proportion of power at different frequencies or polarization may smoothly vary with viewing angle. As another example consider the case where the entries of ρ represent densities of different species, or particles, and allow for the possibility that mass transfers from one species to another, i.e., between entries of ρ. Thus, in general, we postulate that transport of vector-valued quantities captures flow across space as well as between entries of the density vector. We introduce an OMT-inspired geometry that allows us to express a continuity and quantify transport cost for such vectorial distributions.

We begin by considering a vector-valued density ρ on $\mathbb{R}^N$, i.e., a map from $\mathbb{R}^N$ to $\mathbb{R}_+^M$ such that

$$\sum_{i=1}^{M} \int_{\mathbb{R}^N} \rho_i(x)dx = 1.$$

To avoid proliferation of symbols we denote the set of all vector-valued densities and its interior again by $\mathcal{D}$ and $\mathcal{D}_+$, respectively. We refer to the entries of ρ as representing density or mass of species/particles that can mutate between one another while maintaining total mass. The dynamics are captured by the following *continuity equation*:

$$\frac{\partial \rho_i}{\partial t} + \nabla_x \cdot (\rho_i v_i) - \sum_{j \ne i} (\rho_j w_{ji} - \rho_i w_{ij}) = 0 \quad \forall i = 1, \ldots, M. \tag{19}$$

Here v_i is the velocity field of particles i and $w_{ij} \ge 0$ is the transfer rate from i to j. Equation (19) allows for the possibility to mutate between each pair of entries. More generally, mass transfer may only be permissible between specific types of particles and can be modeled by a graph $\mathcal{F} = (\mathcal{V}_1, \mathcal{E}_1, \mathcal{W}_1)$. Thus, (19) corresponds to the case

where $\mathcal{F}$ is a complete graph with all weights equal to 1. For general $\mathcal{F}$ the continuity equation is

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho \circ v) - \nabla^*_{\mathcal{F}}((D_2^T \rho) \circ w - (D_1^T \rho) \circ \bar{w}) = 0. \tag{20}$$

Note here ρ denotes the vector $[\rho_1, \rho_2, \ldots, \rho_M]^T$ and likewise for $v, w, \bar{w}$.

Given $\mu, \nu \in \mathcal{D}_+$, we formulate the optimal mass transport:

$$\begin{aligned} W_{2,b}(\mu,\nu)^2 := \inf_{\rho,v,w,\bar{w}} & \int_0^1 \int_{\mathbb{R}^N} \left\{ v^T(\rho \circ v) + \gamma[w^T((D_2^T\rho)\circ w) \right. \\ & \left. + \bar{w}^T((D_1^T\rho)\circ \bar{w})] \right\} dxdt \\ & \frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho \circ v) - \nabla^*_{\mathcal{F}}((D_2^T \rho) \circ w - (D_1^T \rho) \circ \bar{w}) = 0, \\ & w(t,x) \geq 0, \quad \bar{w}(t,x) \geq 0 \quad \forall t, x \\ & \rho(0,\cdot) = \mu(\cdot), \ \rho(1,\cdot) = \nu(\cdot). \end{aligned} \tag{21}$$

The coefficient $\gamma > 0$ specifies the relative cost between transporting mass in space and trading mass between different types of particles. When γ is large, the solution reduces to independent OMT problems for the different entries to the degree possible. As with $W_{2,a}$, it can be shown that $W_{2,b}$ is a quasi-metric in that it satisfies the triangle inequality and positivity, but is not symmetric. Also, $W_{2,b}$ has the geodesic property

$$W_{2,b}(\rho(s,\cdot), \rho(t,\cdot)) = (t-s) W_{2,b}(\mu,\nu) \tag{22}$$

for $0 \leq s < t \leq 1$, assuming ρ is the optimal flow for (21).

Setting $p = \rho \circ w \geq 0, \bar{p} = \rho \circ \bar{w} \geq 0$ and $u = \rho \circ v$, we establish that (21) is equivalent to the convex optimization problem

$$\begin{aligned} \inf_{\rho,u,p,\bar{p}} & \int_0^1 \int_{\mathbb{R}^N} \left\{ u^T \operatorname{diag}(\rho)^{-1} u + \gamma[p^T \operatorname{diag}(D_2^T\rho)^{-1} p + \bar{p}^T \operatorname{diag}(D_1^T\rho)^{-1}\bar{p}] \right\} dxdt \\ & \frac{\partial \rho}{\partial t} + \nabla_x \cdot u - \nabla^*_{\mathcal{F}}(p - \bar{p}) = 0, \\ & p(t,x) \geq 0, \quad \bar{p}(t,x) \geq 0 \quad \forall t, x \\ & \rho(0,\cdot) = \mu(\cdot), \ \rho(1,\cdot) = \nu(\cdot). \end{aligned} \tag{23}$$

Again, as in (12), one can define a Riemannian like metric on $\mathcal{D}_+$ by symmetrizing the above, which leads to

$$\begin{aligned} \hat{W}_{2,b}(\mu,\nu)^2 & \\ = \inf_{\rho,u,p} & \int_0^1 \int_{\mathbb{R}^N} \left\{ u^T \operatorname{diag}(\rho)^{-1} u + \gamma p^T (\operatorname{diag}(D_2^T\rho)^{-1} + \operatorname{diag}(D_1^T\rho)^{-1}) p \right\} dxdt \\ & \frac{\partial \rho}{\partial t} + \nabla_x \cdot u - \nabla^*_{\mathcal{F}} p = 0, \\ & \rho(0,\cdot) = \mu(\cdot), \ \rho(1,\cdot) = \nu(\cdot). \end{aligned} \tag{24}$$

Remark 5 (Wasserstein-1 distance for vector-valued densities). With continuity equation (20), in the same spirit as in Remark 1, it is straightforward to define the Wasserstein-1 distance for vector-valued densities as

$$W_1(\mu,\nu) = \inf_{u,p} \int \{\|u\| + \gamma\|p\|\}\, dx$$
$$\nu - \mu + \nabla_x \cdot u - \nabla^*_{\mathcal{F}}\, p = 0,$$

whose dual is clearly

$$W_1(\mu,\nu) = \sup_{f} \int \{f^T(\nu-\mu)\} dx$$
$$\|\nabla_x f\| \le 1, \quad \|\nabla_{\mathcal{F}} f\| \le \gamma.$$

5. Vector-valued mass transport on graphs. We finally consider vector-valued mass transport on graphs. A vector-valued mass distribution on graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ (with n nodes and m edges) is an M-tuple $\rho = (\rho_1, \dots, \rho_M)$ with each $\rho_i = (\rho_{i,1}, \dots, \rho_{i,n})^T$ being a vector in $\mathbb{R}^n_+$ such that

$$\sum_{i=1}^{M} \sum_{k=1}^{n} \rho_{i,k} = 1.$$

That is, each entry ρ_i, for $i \in \{1, \dots, M\}$, is a vector with nonnegative n-entries representing, e.g., color intensity for the ith color, at the node corresponding to the respective entry. We denote the set of all nonnegative vector-valued mass distributions with $\mathcal{D}$ and its interior with $\mathcal{D}_+$. Combining (7) and (19) we obtain the continuity equation

$$\dot\rho - \nabla^*_{\mathcal{G}}((D_2^T\rho)\circ v - (D_1^T\rho)\circ \bar v) - \nabla^*_{\mathcal{F}}((D_2^T\rho)\circ w - (D_1^T\rho)\circ \bar w) = 0. \tag{25}$$

The problem of transporting vector-valued mass on a graph is conceptually simpler as it reduces essentially to a scalar mass situtation. Indeed, we can view the vector-valued mass as a scalar mass distribution on M identical layers of the graph $\mathcal{G}$ where the same nodes at different layers are connected through a graph $\mathcal{F}$. The two velocity fields v, w represent mass transfer within the same layer and between different layers, respectively.

Following our earlier program, given two marginal densities $\mu, \nu \in \mathcal{D}_+$, we define their Wasserstein distance as

$$\begin{aligned} W_{2,c}(\mu,\nu)^2 := \inf_{\rho,v,\bar v,w,\bar w} \int_0^1 & \left\{ v^T((D_2^T\rho)\circ v) + \bar v^T((D_1^T\rho)\circ \bar v) \right. \\ & \left. + \gamma\left[w^T((D_2^T\rho)\circ w) + \bar w^T((D_1^T\rho)\circ \bar w)\right]\right\} dt \\ & \dot\rho - \nabla^*_{\mathcal{G}}((D_2^T\rho)\circ v - (D_1^T\rho)\circ \bar v) - \nabla^*_{\mathcal{F}}((D_2^T\rho)\circ w - (D_1^T\rho)\circ \bar w) = 0, \\ & w \ge 0,\ \bar w \ge 0,\ v \ge 0,\ \bar v \ge 0, \\ & \rho(0) = \mu,\ \rho(1) = \nu. \end{aligned}$$

In the same way as before, the above has a convex reformulation,

$$\begin{aligned}
&\inf_{\rho,u,\bar{u},p,\bar{p}} \int_0^1 \left\{ u^T \operatorname{diag}(D_2^T \rho)^{-1} u + \bar{u}^T \operatorname{diag}(D_1^T \rho)^{-1} \bar{u} \right. \\
&\left. + \gamma [p^T \operatorname{diag}(D_2^T \rho)^{-1} p + \bar{p}^T \operatorname{diag}(D_1^T \rho)^{-1} \bar{p}] \right\} dt \\
&\dot{\rho} - \nabla_{\mathcal{G}}^* (u - \bar{u}) - \nabla_{\mathcal{F}}^* (p - \bar{p}) = 0, \\
&p \geq 0,\ \bar{p} \geq 0,\ u \geq 0,\ \bar{u} \geq 0, \\
&\rho(0) = \mu,\ \rho(1) = \nu.
\end{aligned}$$

The same method as in (12) gives rise to a symmetric Riemannian type metric $\hat{W}_{2,c}(\mu,\nu)^2$ provided by the solution of

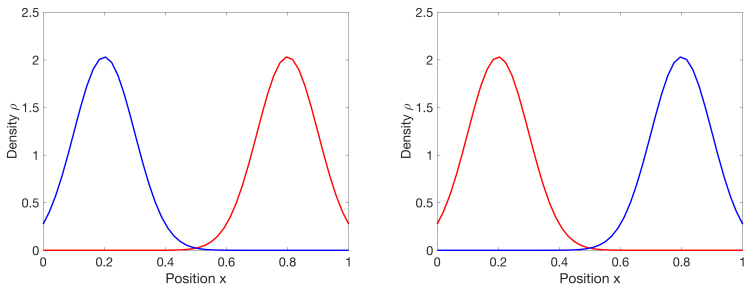
$$\begin{aligned}
&\inf_{\rho,u,p} \int_0^1 \left\{ u^T (\operatorname{diag}(D_2^T \rho)^{-1} + \operatorname{diag}(D_1^T \rho)^{-1}) u \right. \\
&\left. + \gamma p^T \ (\operatorname{diag}(D_2^T \rho)^{-1} + \operatorname{diag}(D_1^T \rho)^{-1}) p \right\} dt \ \dot{\rho} - \nabla_{\mathcal{G}}^* u - \nabla_{\mathcal{F}}^* p = 0, \\
&\rho(0) = \mu,\ \rho(1) = \nu.
\end{aligned}$$

6. Examples. In this section, we present two examples. The first one is an academic example to illustrate the idea of vector-valued optimal mass transport. In the second example, we apply our framework to color image processing problems.

6.1. Interpolation of one-dimensional densities. We consider vector-valued densities with two components on the real line (interval [0, 1]). The two marginal densities μ and ν are displayed in Figure 1 with the two colors (red and blue) denoting the two components.

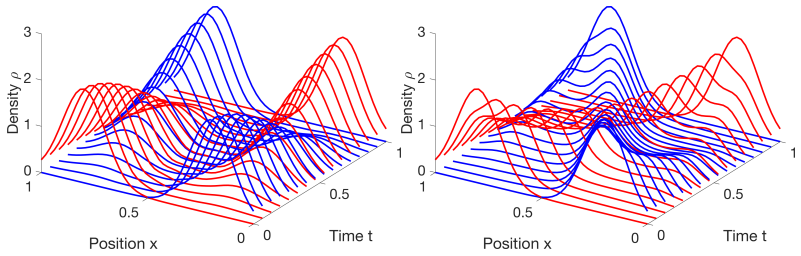
We solve the symmetric vector-valued transport problem (24) for several different values of γ. For the numerical implementation, we first discretize the space interval [0, 1] to convert it into a vector-valued transport problems on graphs, which is essentially (12). Then we discretize the time dimension with staggered grids. In particular, we discretize the time interval into n_t subintervals. Then the densities take value at time points $0, 1/n_t, \ldots, (n_t - 1)/n_t, 1$ while the fluxes take values at time points $1/(2n_t), 3/(2n_t), \ldots, (2n_t - 1)/(2n_t)$. We refer the reader to [7] for more details about the convex optimization algorithm used for the examples in this paper.

The results are depicted in Figure 2. As can be seen, for large γ, the solution tends to have two independent transport plans, as the cost of transferring between the two different masses is high. In contrast, when γ is small, the solution prefers transferring rather than transporting, since the cost of transferring between masses is low.

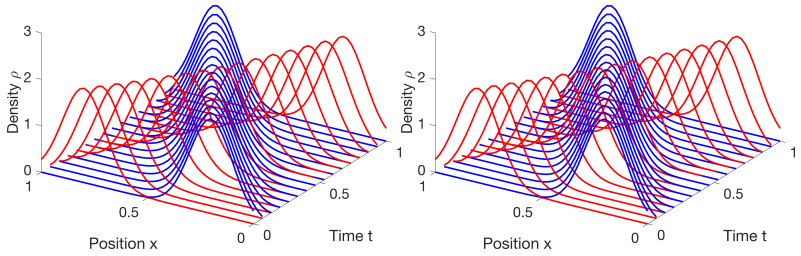


(a) μ (b) ν

FIG. 1. *Marginal distributions.*

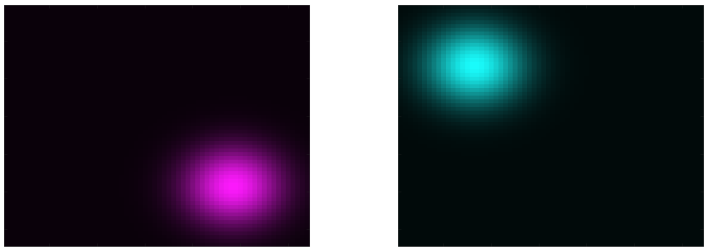


(a) $\gamma = 0.0001$ (b) $\gamma = 0.01$



(c) $\gamma = 1$ (d) $\gamma = 100$

FIG. 2. *Density flow with different γ values.*

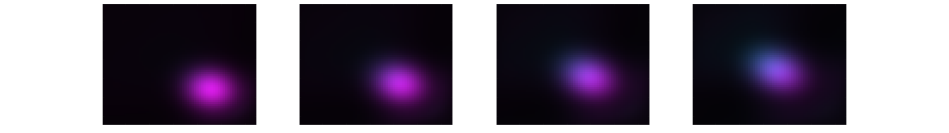


(a) ρ_0 (b) ρ_1

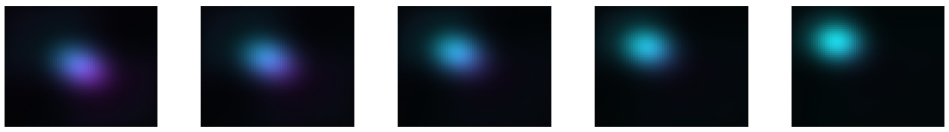
FIG. 3. *Marginal distributions.*

6.2. Interpolation of color images. As alluded to previously, the different components of a image may stand for different color channels. For instance, a color image can be viewed as a vector-valued density with three components that represent red (R), green (G), blue (B), respectively. Thus, it is straightforward to use vector-valued optimal mass transport to compare and interpolate such color images. Below we explain three representative examples, shown in Figures 3 through 8, that highlight the mechanism of vector-valued transport.

First consider the two color images (64×64) shown in Figure 3. The intensity in each is a Gaussian distribution centered at a different location. The two distributions are of different color, thereby the corresponding vectorial-valued mass is distributed differently across the three components. The result of interpolating between the two, with $\gamma = 0.001$, is shown in Figure 4. As we can see from the subplots, the displacement of the mass appears to run at constant speed between ρ_0 to ρ_1 while, at the same time, the color is changing gradually as mass flows between the vectorial components.

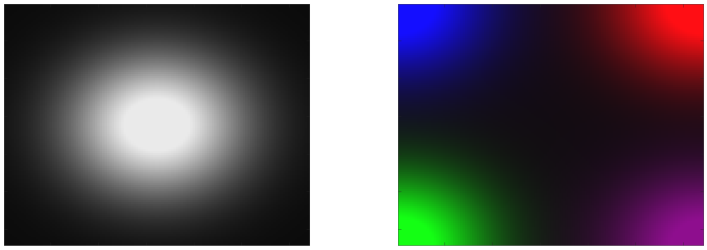


(a) $t = 0.1$ (b) $t = 0.2$ (c) $t = 0.3$ (d) $t = 0.4$



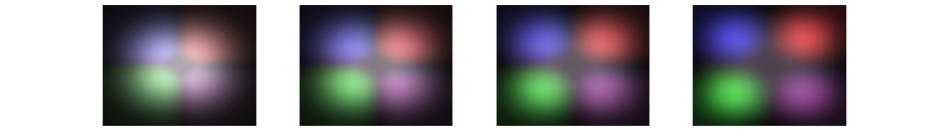
(e) $t = 0.5$ (f) $t = 0.6$ (g) $t = 0.7$ (h) $t = 0.8$ (i) $t = 0.9$

FIG. 4. *Interpolation with* $\gamma = 0.001$.

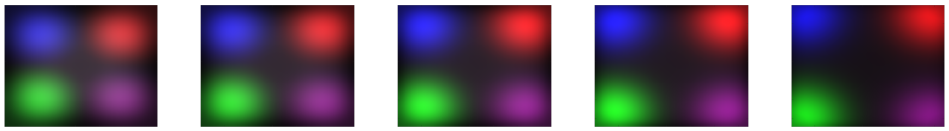


(a) ρ_0 (b) ρ_1

FIG. 5. *Marginal distributions.*



(a) $t = 0.1$ (b) $t = 0.2$ (c) $t = 0.3$ (d) $t = 0.4$

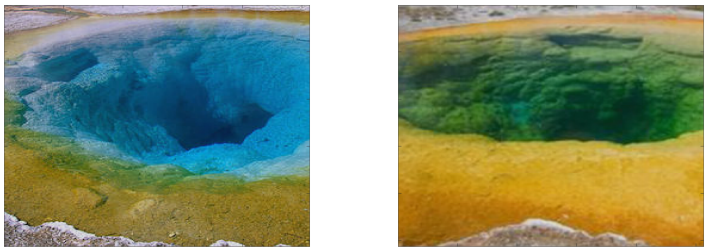


(e) $t = 0.5$ (f) $t = 0.6$ (g) $t = 0.7$ (h) $t = 0.8$ (i) $t = 0.9$

FIG. 6. *Interpolation with* $\gamma = 0.01$.

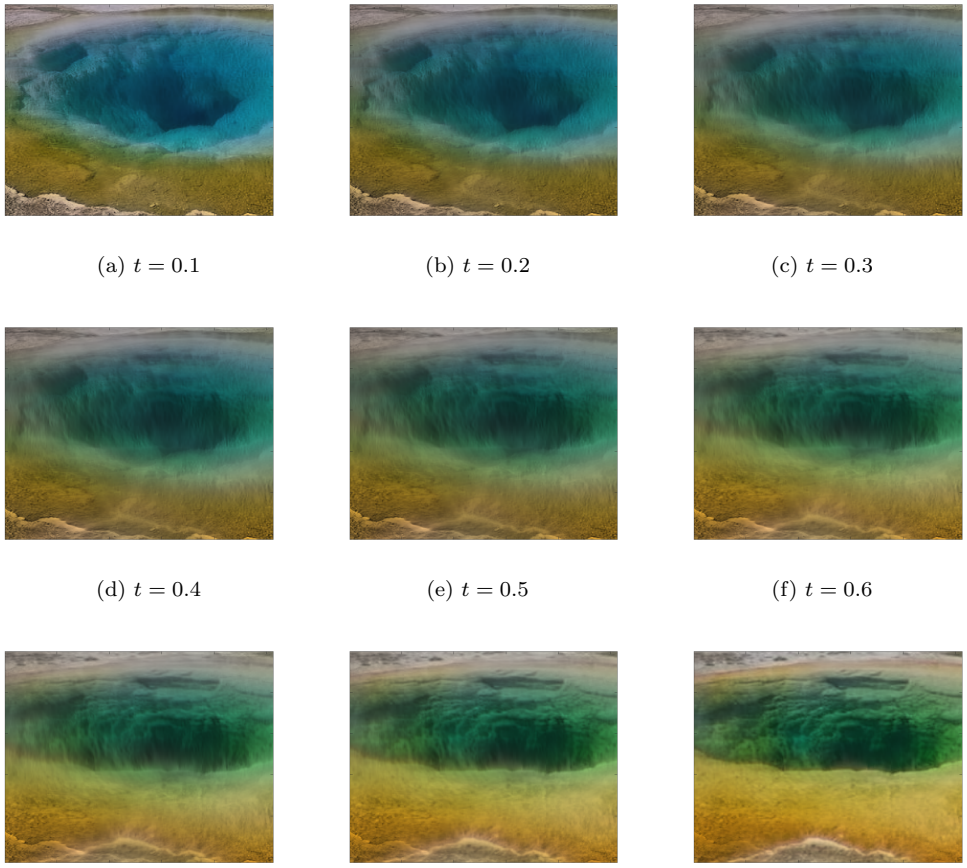
Figure 5 shows yet another example of a similar nature. The initial density is centered and it is white, which signifies equal mass distribution across the three color channels/components. The terminal density on the other hand has four separated masses of different color. The dimensions of the images are 128 by 128. The density flow shown in Figure 6, based on our technique with $\gamma = 0.01$, smoothly interpolates by dispacing the intensity and color profiles in a seemingly natural manner.

Finally, in Figures 7 and 8 we display the result of interpolating real-life images. The marginal distributions shown in Figure 7 are two photos (256×256) of two geothermal basins in Yellowstone Park, where bacterial growth give them distinctly different colors and hues. The result of interpolating the corresponding vector-valued distributions is depicted in Figure 8, taking $\gamma = 0.3$. The flow of images produces a sequence of natural looking images transitioning from one to the next.



(a) ρ_0 (b) ρ_1

FIG. 7. *Marginal distributions.*



(a) $t = 0.1$ (b) $t = 0.2$ (c) $t = 0.3$

(d) $t = 0.4$ (e) $t = 0.5$ (f) $t = 0.6$

(g) $t = 0.7$ (h) $t = 0.8$ (i) $t = 0.9$

FIG. 8. *Interpolation with* $\gamma = 0.3$.

In all the examples, we observe the apparently natural displacement of intensity and color that should be contrasted with potentially undesirable "push-pop" effects of linear interpolation.

7. Conclusions and further research. Our early motivation, as noted in the Introduction, has been to devise a suitable geometry to study flows of probability or power distribution in problems of signal analysis and fusion of vectorial data. However, the framework, very much as the broader subject of optimal mass transport, has application in a wider range of ideas. In particular, the connection between transport geometry and properties of an underlying space (e.g., curvature in the Bakry–Emery theory) may have important implications here as well. More specifically, we are interested in applying this methodology to studying the robustness of various networks, as was done in [24, 25] for biological and financial networks and in [30] for communications networks.

7.1. Biological networks. The study of cellular networks (e.g., signaling and transcription) has become a major enterprise in systems biology; see [1] and the references therein. One of the key problems is understanding global properties of cellular networks, in order to differentiate a diseased state from a normal cellular state. As is argued in several places [10, 31, 24], network properties may help in formulating systems biological concepts that could lead to novel therapies for a number of diseases, including cancer. This would involve integrating genetic, epigenetic, and protein-protein interaction networks.

7.2. Financial networks. Stock data and financial transactions provide an insight into the vast global financial network of human activities. The health of the national and world economy is reflected in the robustness and self-regulatory properties of the markets. Long range correlations are responsible for cascade failures due to financial insolvency. Indeed, multiple exposures of companies is often the root cause of infectious propagation of balance sheet insolvency, with catastrophic effects. It is of interest to understand the relation between the various financial parameters (assets, liability, capital) that quantify the stress and the buffer capabilities of financial institutions with network connectivity and interdependence (weighted network Laplacian) so as to assess risk of cascade failures and fragility, and to devise ways to mitigate such effects. See [25] and the references therein.

At closer inspection, many of the aforementioned problem areas involve finer attributes of the studied objects, which may be more suitably treated and studied as vector-valued distributions. Thus, we hope that the present work provides a starting point for such an endeavor.

REFERENCES

[1] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, CRC Press, Boca Raton, FL, 2006.

[2] D. Bakry and M. Émery, *Diffusions hypercontractives*, in Séminaire de Probabilités, XIX, Lect. Notes Math. 1123 (Springer, Berlin, 1985), pp. 177–206.

[3] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[4] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum, *On the matrix Monge-Kantorovich problem*, preprint, arXiv:1701.02826, 2017.

[5] 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., 169 (2016), pp. 671–691.

[6] Y. Chen, T. T. Georgiou, and A. Tannenbaum, *Matrix optimal mass transport: A quantum mechanical approach*, IEEE Trans. Automat. Control, to appear.
[7] Y. Chen, K. Yamamoto, E. Haber, T. T. Georgiou, and A. Tannenbaum, *An efficient algorithm for matrix-valued and vector-valued optimal mass transport*, J. Sci. Comput., to appear.
[8] S.-N. Chow, W. Huang, Y. Li, and H. Zhou, *Fokker–Planck equations for a free energy functional or Markov process on a graph*, Arch. Ration. Mech. Anal., 203 (2012), pp. 969–1008.
[9] S.-N. Chow, W. Li, and H. Zhou, *Entropy dissipation of Fokker–Planck equations on graphs*, preprint, arXiv:1701.04841, 2017.
[10] L. Demetrius and T. Manke, *Robustness and network evolution – An entropic principle*, Physica A, 346 (2005), pp. 682–696.
[11] M. Erbar and J. Maas, *Ricci curvature of finite Markov chains via convexity of the entropy*, Arch. Ration. Mech. Anal., 206 (2012), pp. 1–42.
[12] L. C. Evans and W. Gangbo, *Differential Equations Methods for the Monge-Kantorovich Mass Transfer Problem*, Mem. Amer. Math. Soc. 653, AMS, Providence, 1999.
[13] T. T. Georgiou, J. Karlsson, and M. S. Takyar, *Metrics for power spectra: An axiomatic approach*, IEEE Trans. Signal Process., 57 (2009), pp. 859–867.
[14] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, *Optimal mass transport for registration and warping*, Int. J. Comput. Vis., 60 (2004), pp. 225–240.
[15] X. Jiang, L. Ning, and T. T. Georgiou, *Distances and Riemannian metrics for multivariate spectral densities*, IEEE Trans. Automat. Control, 57 (2012), pp. 1723–1735.
[16] L. V. Kantorovich, *On a problem of Monge*, J. Math. Sci., 3 (1948), pp. 225–226.
[17] C. Léonard, *Lazy random walks and optimal transport on graphs*, Ann. Probab., 44 (2016), pp. 1864–1915.
[18] W. Li, P. Yin, and S. Osher, *A fast algorithm for unbalanced L1 Monge-Kantorovich problem*, CAM Rep., 2016.
[19] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. Math., (2009), pp. 903–991.
[20] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.
[21] L. Ning, T. T. Georgiou, and A. Tannenbaum, *On matrix-valued Monge-Kantorovich optimal mass transport*, IEEE Trans. Automat. Control, 60 (2015), pp. 373–382.
[22] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation*, Commun. Partial Differ. Equ., 26 (2001), pp. 101–174.
[23] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*, Springer, Berlin, 1998.
[24] R. Sandhu, T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum, *Graph curvature for differentiating cancer networks*, Sci. Rep., 5 (2015), p. 12323.
[25] R. S. Sandhu, T. T. Georgiou, and A. R. Tannenbaum, *Ricci curvature: An economic indicator for market fragility and systemic risk*, Sci. Adv., 2 (2016), p. e1501495.
[26] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, *Continuous-flow graph transportation distances*, preprint, arXiv:1603.06927, 2016.
[27] E. Tannenbaum, T. Georgiou, and A. Tannenbaum, *Signals and control aspects of optimal mass transport and the Boltzmann entropy*, in Proceedings of the 2010 49th IEEE Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2010, pp. 1885–1890.
[28] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, AMS, Providence, 2003.
[29] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Commun. Pure Appl. Math., 58 (2005), pp. 923–940.
[30] C. Wang, E. Jonckheere, and R. Banirazi, *Wireless network capacity versus Ollivier-Ricci curvature under heat-diffusion (HD) protocol*, in American Control Conference (ACC), 2014, IEEE, Piscataway, NJ, 2014, pp. 3536–3541.
[31] J. West, G. Bianconi, S. Severini, and A. E. Teschendorff, *Differential network entropy reveals cancer system hallmarks*, Sci. Rep., 2 (2012), p. 802.