Published in final edited form as:

IEEE Trans Automat Contr. 2015 February ; 60(2): 373–382. doi:10.1109/TAC.2014.2350171.

On Matrix-Valued Monge–Kantorovich Optimal Mass Transport

Lipeng Ning,

Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115 USA

Tryphon T. Georgiou, Fellow, IEEE, and

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA

Allen Tannenbaum, Fellow, IEEE

Departments of Computer Science and Applied Mathematics, Stony Brook University, Stony Brook, NY 11794 USA

Lipeng Ning: lning@bwh.harvard.edu; Tryphon T. Georgiou: tryphon@umn.edu; Allen Tannenbaum: allen.tannenbaum@stonybrook.edu

Abstract

We present a particular formulation of optimal transport for matrix-valued density functions. Our aim is to devise a geometry which is suitable for comparing power spectral densities of multivariable time series. More specifically, the value of a power spectral density at a given frequency, which in the matricial case encodes power as well as directionality, is thought of as a proxy for a “matrix-valued mass density.” Optimal transport aims at establishing a natural metric in the space of such matrix-valued densities which takes into account differences between power across frequencies as well as misalignment of the corresponding principle axes. Thus, our transportation cost includes a cost of transference of power between frequencies together with a cost of rotating the principle directions of matrix densities. The two endpoint matrix-valued densities can be thought of as marginals of a joint matrix-valued density on a tensor product space. This joint density, very much as in the classical Monge–Kantorovich setting, can be thought to specify the transportation plan. Contrary to the classical setting, the optimal transport plan for matrices is no longer supported on a thin zero-measure set.

Index Terms

Convex optimization; matrix-valued density functions; optimal mass-transport

I. Introduction

The problem of optimal mass transport (OMT) dates back to the work of G. Monge in 1781 [1] while its modern formulation is due to L. Kantorovich [2]. In recent years the subject is developing rapidly due to its intrinsic significance and range of applications in physics, economics, and probability [3]–[5].

Our motivation for studying “matrix-valued transport” originates in the spectral analysis of multi-variable time series. Just as in scalar time series, spectral content is assessed based on (estimated) statistics of the underlying process, where these are simply moments of the corresponding power spectral density (PSD). Different metrics have been proposed to compare PSD’s for purposes of spectral approximation, estimation, and system modeling (see [6], [7] and the references therein). However, since spectra are estimated based on integrals, weak* metrics¹ are preferable since they provide continuity of statistics to perturbations in the PSD. Earlier metrics and so called, divergence measures, typically fail in this respect (see [7]). Hence, for this reason, optimal mass transport which endows the space of (scalar) probability/mass/power densities with a natural weak* metric—the Wasserstein metric, is of particular interest. Our aim in this paper is to develop one possible such generalization of the Wasserstein metric that allows comparison of matrix-valued density functions in a similar spirit.

The scalar OMT theory has been adapted in [8] to model slowly time-varying changes in power spectra of time series and has been used for statistical estimation, data assimilation, and morphing. While in scalar time series, the power spectral content may drift across frequencies over time (e.g., when considering Doppler effects, echolocation of a moving target, etc.), in vector-valued time series the power spectral content may shift principle directions as well. In fact, such a rotation of the power-spectral content is typical in general antenna-arrays when a scatterer changes position with respect to array elements. Therefore, a concept of transport between matrix-valued densities requires that we take into account both, the cost of shifting power across frequencies as well as the cost of rotating the corresponding principle axes. Besides our particular formulation of a “non-commutative” Monge–Kantorovich transportation problem and of a corresponding metric, the main results in this paper are i) that the optimal transport can be cast as a convex-optimization problem, ii) the geodesics and transport paths can be determined using convex programming, and iii) the optimal transport plan has support which, in contrast to the classical Monge–Kantorovich setting, is no longer contained on a thin zero-measure set. The relevance of the proposed metric is highlighted in examples on spectral morphing and spectral tracking in the final section of the paper.

II. Preliminaries on Optimal Mass Transport

Consider two probability density functions μ_0 and μ_1 supported on $\mathbb{R}$ and let $\mathcal{M}(\mu_0, \mu_1)$ be the set of probability measures m on $\mathbb{R} \times \mathbb{R}$ with μ_0 and μ_1 as marginals, i.e.,

$$\int_{\mathbb{R}} m(x, y) dy = \mu_0(x), \quad \int_{\mathbb{R}} m(x, y) dx = \mu_1(y), \quad m(x, y) \geq 0.$$

Clearly, $\mathcal{M}(\mu_0, \mu_1)$ is not empty since already the product $\mu_0(x)\mu_1(y) \in \mathcal{M}(\mu_0, \mu_1)$. Probability densities are thought of as distributions of mass and the optimal mass transport problem is to determine

¹A sequence of measures $d\mu_n$ converges weak* to $d\mu$ if and only if $\int f d\mu_n \rightarrow \int f d\mu$ for all continuous and bounded f .

$$\mathcal{T}_c(\mu_0, \mu_1) := \inf_{m \in \mathcal{M}(\mu_0, \mu_1)_{\mathbb{R} \times \mathbb{R}}} \int c(x, y) m(x, y) dx dy \quad (1)$$

where $c(x, y)$ is the cost of transporting one unit of mass from location x to y . In particular, when $c(x, y) = |x - y|^2$, the optimal cost gives rise to the 2-Wasserstein metric

$$W_2(\mu_0, \mu_1) = \mathcal{T}_2(\mu_0, \mu_1)^{\frac{1}{2}}$$

where

$$\mathcal{T}_2(\mu_0, \mu_1) := \inf_{m \in \mathcal{M}(\mu_0, \mu_1)_{\mathbb{R} \times \mathbb{R}}} \int |x - y|^2 m(x, y) dx dy. \quad (2)$$

In general, (1) is a linear program with dual

$$\sup_{\phi, \psi} \left\{ \int_{\mathbb{R}} (\phi_0(x) \mu_0(x) - \phi_1(x) \mu_1(x)) dx \mid \phi_0(x) - \phi_1(y) \leq c(x, y) \right\} \quad (3)$$

where ϕ_0, ϕ_1 are continuous, see [3]. For the quadratic cost $c(x, y) = |x - y|^2$ and in one spatial dimension, $\mathcal{T}_2(\mu_0, \mu_1)$ can also be written explicitly in terms of the cumulative distributions functions

$$M_i(x) = \int_{-\infty}^x \mu_i(x) dx \quad \text{for } i=0, 1$$

in the form

$$\mathcal{T}_2(\mu_0, \mu_1) = \int_0^1 |M_0^{-1}(t) - M_1^{-1}(t)|^2 dt. \quad (4)$$

In this case, the optimal joint probability density $m \in \mathcal{M}(\mu_0, \mu_1)$ has support on $(x, T(x))$ where $T(x)$ is the sub-differential of a convex lower semi-continuous function, see [3, p. 75]. More specifically, $T(x)$ is uniquely defined by

$$M_0(x) = M_1(T(x)). \quad (5)$$

Interestingly, a geodesic μ_τ ($\tau \in [0, 1]$) between μ_0 and μ_1 can be written explicitly as well in terms of a corresponding cumulative function M_τ for each τ , defined via

$$M_\tau((1-\tau)x + \tau T(x)) = M_0(x). \quad (6)$$

Indeed, it readily follows that:

$$\begin{aligned} W_2(\mu_0, \mu_\tau) &= \tau W_2(\mu_0, \mu_1) \\ W_2(\mu_\tau, \mu_1) &= (1-\tau) W_2(\mu_0, \mu_1) \end{aligned}$$

and that μ_τ ($\tau \in [0, 1]$) is a geodesic.

III. Matrix-Valued Optimal Mass Transport

We consider the one-dimensional family of matrix-valued functions

$$\mathcal{F} := \left\{ \mu(\cdot) \mid \text{for } x \in \mathbb{R}, \mu(x)^* = \mu(x) \in \mathbb{C}^{n \times n}, \mu(x) \geq 0, \text{tr} \left(\int_{\mathbb{R}} \mu(x) dx \right) = 1 \right\}.$$

These are Hermitian, positive semi-definite matrix-valued functions on $\mathbb{R}$ normalized so that their trace integrates to 1. They will be referred to as matrix-valued densities and can be thought of as a generalization of probability density functions. The scalar-valued $\text{tr}(\mu)$ represents mass at location x . Thus, all elements in $\mathcal{F}$ have the same total mass over the support. Below, we motivate a particular cost of transportation between such matrix-valued functions and introduce a suitable generalization of the Monge–Kantorovich OMT to matrix-valued densities.

A. Tensor Product and Partial Trace

Consider two n -dimensional real or complex (Hilbert) spaces $\mathcal{H}_0$ and $\mathcal{H}_1$, let $\mathcal{L}(\mathcal{H}_0)$ and $\mathcal{L}(\mathcal{H}_1)$ denote the space of linear operators on $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, and let $\mu_0 \in \mathcal{L}(\mathcal{H}_0)$ and $\mu_1 \in \mathcal{L}(\mathcal{H}_1)$. Thus, in the present subsection, μ_i ($i \in \{0, 1\}$) are fixed matrices. We denote their tensor product by $\mu_0 \otimes \mu_1 \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ which is formally defined via

$$\mu_0 \otimes \mu_1 : u \otimes v \mapsto \mu_0 u \otimes \mu_1 v.$$

Since our spaces are finite-dimensional this can be identified with the Kronecker product of the corresponding matrix representation of the two operators. The space $\mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ is the span of all products $\mu_0 \otimes \mu_1$ with $\mu_i \in \mathcal{L}(\mathcal{H}_i)$ for $i \in \{0, 1\}$.

Consider $\mu \in \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1)$. The partial traces $\text{tr}_{\mathcal{H}_0}$ and $\text{tr}_{\mathcal{H}_1}$, or tr_0 and tr_1 for brevity, are linear maps

$$\begin{aligned} \text{tr}_1 : \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1) &\rightarrow \mathcal{L}(\mathcal{H}_0) : \mu \mapsto \text{tr}_1(\mu) \\ \text{tr}_0 : \mathcal{L}(\mathcal{H}_0 \otimes \mathcal{H}_1) &\rightarrow \mathcal{L}(\mathcal{H}_1) : \mu \mapsto \text{tr}_0(\mu) \end{aligned}$$

defined uniquely by the property that on simple products they act as follows:

$$\text{tr}_1(\mu_0 \otimes \mu_1) = \text{tr}(\mu_1) \mu_0 \quad \text{and} \quad \text{tr}_0(\mu_0 \otimes \mu_1) = \text{tr}(\mu_0) \mu_1$$

for any $\mu_0 \in \mathcal{D}(\mathcal{H}_0)$ and $\mu_1 \in \mathcal{D}(\mathcal{H}_1)$. Alternatively, $\mu \in \mathcal{D}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ can be represented by a matrix $[\mu_{ik, \ell m}]$ of size $n^2 \times n^2$ as it maps a basis element $u_i \otimes v_k \in \mathcal{H}_0 \otimes \mathcal{H}_1$ to $\sum_{\ell, m} \mu_{ik, \ell m} u_\ell \otimes v_m$. Then, the partial trace e.g., $\text{tr}_1(\mu)$ is represented by the $n \times n$ matrix with (i, ℓ) -th entry $\sum_k \mu_{ik, \ell k}$, for $1 \leq i, \ell \leq n$. Likewise the (k, m) -th entry of $\text{tr}_0(\mu)$ is $\sum_i \mu_{ik, im}$, for $1 \leq k, m \leq n$. See [9] for the significance of partial trace in the context of quantum mechanics.

B. Joint Matrix-Valued Density

We now return to considering matrix-valued density functions $\mu_0, \mu_1 \in \mathcal{F}$. A naive attempt is to seek a joint density $m = 0$ having support on $\mathbb{R} \times \mathbb{R}$ and having μ_0, μ_1 as “marginals,” i.e., so that

$$\int_{\mathbb{R}} m(x, y) dy = \mu_0(x), \quad \int_{\mathbb{R}} m(x, y) dx = \mu_1(y). \quad (7)$$

However, in contrast to the scalar case, such an m does not exist in general. To see this, consider the case of matrix valued measures

$$\begin{aligned} \mu_0(x) &= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} \delta(x-1) + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \delta(x-2), \\ \mu_1(y) &= \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix} \delta(y-1) + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \delta(y-2) \end{aligned}$$

where $\delta(\cdot)$ denotes the Dirac delta. If $m(x, y)$ were to exist, its support would be contained in $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$. It is easy to see that there cannot be a consistent selection of four 2×2 matrices so that, in pairs, they sum up to the coefficients making up $\mu_0(x)$ and $\mu_1(y)$.

Thus, any natural definition of a transportation plan requires that the joint density lives in a bigger space. A particular formulation is as follows: we seek

$$m(x, y) \text{ being } n^2 \times n^2 \text{ positive semi-definite matrix} \quad (8a)$$

for $(x, y) \in \mathbb{R} \times \mathbb{R}$, such that

$$m_0(x, y) := \text{tr}_1(m(x, y)), \quad m_1(x, y) := \text{tr}_0(m(x, y)) \quad (8b)$$

$$\int_{\mathbb{R}} m_0(x, y) dy = \mu_0(x), \quad \int_{\mathbb{R}} m_1(x, y) dx = \mu_1(y) \quad (8c)$$

and denote

$$M(\mu_0, \mu_1) := \{m \mid (8a) - (8c) \text{ are satisfied}\}.$$

Since $\mu_0 \otimes \mu_1 \in \mathcal{M}(\mu_0, \mu_1)$, the set $\mathcal{M}(\mu_0, \mu_1)$ is clearly not empty.

We next motivate a suitable transportation cost. This is a functional on the joint density $\mathcal{M}(\mu_0, \mu_1)$, just as in the scalar case. However, besides penalizing transport of mass between two points x and y , we also impose a penalty on a corresponding rotation as well.

C. Transportation Cost

As indicated earlier, we interpret $\text{tr}(\mathbf{m}(x, y))$ as the total “mass” that is being transferred from x to y . We consider a scalar cost² $c(x, y) = (x - y)^2$ and the “mass transference” cost

$$\min_{\mathbf{m} \in \mathcal{M}(\mu_0, \mu_1)_{\mathbb{R} \times \mathbb{R}}} \int c(x, y) \text{tr}(\mathbf{m}(x, y)) dx dy. \quad (9)$$

This coincides with the optimal transportation cost between scalar-valued densities $\text{tr}(\mu_0)$ and $\text{tr}(\mu_1)$. Thus, if $\text{tr}(\mu_0(x)) = \text{tr}(\mu_1(x))$, the optimal value of (9) is zero since it reduces to optimal transport between identical scalar marginals. Thus, (9) fails to quantify mismatch of directionality between the given matrix-valued marginals. Below, we introduce a term that penalizes directionality mismatch.

We assume throughout that the marginals are positive definite pointwise. Then, for $i \in \{0, 1\}$, $\text{tr}(\mu_i(x))$ represents the total mass at x while $\mu_i(x)/\text{tr}(\mu_i(x))$, normalized to have trace 1, encapsulates directional information. Likewise, for the joint density $\mathbf{m}(x, y)$, assuming that $\mathbf{m}(x, y) \succeq 0$, we define the *normalized partial traces*

$$\begin{aligned} \underline{\text{tr}}_0(\mathbf{m}(x, y)) &:= \text{tr}_0(\mathbf{m}(x, y)) / \text{tr}(\mathbf{m}(x, y)) \\ \underline{\text{tr}}_1(\mathbf{m}(x, y)) &:= \text{tr}_1(\mathbf{m}(x, y)) / \text{tr}(\mathbf{m}(x, y)). \end{aligned}$$

Their difference captures the directional mismatch between the two partial traces. Hence, we introduce

$$\text{tr} \left(\|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1)\mathbf{m}(x, y)\|_{\text{F}}^2 \mathbf{m}(x, y) \right)$$

to quantify the rotational mismatch and we consider the cost functional

$$\text{tr} \left(\left(c(x, y) + \lambda \|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1)\mathbf{m}(x, y)\|_{\text{F}}^2 \right) \mathbf{m}(x, y) \right)$$

with $\lambda > 0$, to weigh in the relative significance of the linear and rotational penalties.

²It is interesting to consider functionals of the form $\text{tr}(\mathbf{c}(x, y)\mathbf{m}(x, y))$, with $\mathbf{c}(x, y)$ being matricial, and how to utilize such so as to reflect transportation cost with practical relevance.

D. Optimal Transportation Problem

In view of the above, we define

$$\mathcal{T}_{2,\lambda}(\mu_0, \mu_1) := \min_{\mathbf{m} \in \mathcal{M}(\mu_0, \mu_1)_{\mathbb{R} \times \mathbb{R}}} \int \text{tr} \left((c + \lambda \|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1) \mathbf{m}\|_{\text{F}}^2) \mathbf{m} \right) dx dy \quad (10)$$

with $c(x, y) = (x - y)^2$, and show next that (10) is in fact a convex optimization problem.

From the definition

$$\begin{aligned} \underline{\text{tr}}_0(\mathbf{m}) \text{tr}(\mathbf{m}) &= \text{tr}_0(\mathbf{m}), \\ \underline{\text{tr}}_1(\mathbf{m}) \text{tr}(\mathbf{m}) &= \text{tr}_1(\mathbf{m}) \end{aligned}$$

and hence

$$\begin{aligned} \|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1) \mathbf{m}\|_{\text{F}}^2 \text{tr}(\mathbf{m}) &= \frac{\|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1) \mathbf{m}\|_{\text{F}}^2 \text{tr}(\mathbf{m})^2}{\text{tr}(\mathbf{m})} \\ &= \frac{\|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1) \mathbf{m}\|_{\text{F}}^2}{\text{tr}(\mathbf{m})}. \end{aligned}$$

Now let $m(x, y) = \text{tr}(\mathbf{m}(x, y))$ and $\mathbf{m}_0(x, y)$, $\mathbf{m}_1(x, y)$ be as in (8). The expression for the optimal cost in (10) is lower bounded by

$$\begin{aligned} \min_{\mathbf{m}_0, \mathbf{m}_1, m} \left\{ \int \left(c(x, y) m(x, y) + \lambda \frac{\|\mathbf{m}_0 - \mathbf{m}_1\|_{\text{F}}^2}{m} \right) dx dy \right. \\ \left. \begin{aligned} &\mathbf{m}_0(x, y), \mathbf{m}_1(x, y) \geq 0, \\ &\text{tr}(\mathbf{m}_0(x, y)) = \text{tr}(\mathbf{m}_1(x, y)) = m(x, y) \\ &\int \mathbf{m}_0(x, y) dy = \mu_0(x), \int \mathbf{m}_1(x, y) dx = \mu_1(y) \end{aligned} \right\}. \end{aligned} \quad (11)$$

For an optimal triple $m, \hat{\mathbf{m}}_0, \hat{\mathbf{m}}_1$ of (11), $\mathbf{m} := \hat{\mathbf{m}}_0 \otimes \hat{\mathbf{m}}_1$ is a minimizer of (10) that gives the same optimal value as (11). Thus, the optimal cost in (10) is equivalently written as (11).

For $x > 0$, the expression $(y - z)^2/x$ is jointly convex in the arguments x, y, z , see e.g., [10, p. 72]. It readily follows that the integral in (11) is a jointly convex functional of its arguments. All additional constraints in (11) are convex as well and, therefore, so is the optimization problem.

IV. On the Geometry of Optimal Mass Transport

An important result in the (scalar) OMT theory is that the transportation plan is the sub-differential of a convex function and has support on a thin zero-measure set, see e.g., [3, p. 92]. This property is not shared by the optimal transportation plan between matrix-valued density functions as we explain next.

In standard scalar OMT with convex transportation cost, the optimal transportation plan has a certain cyclically monotonic property [3]. More specifically, if $(x_1, y_1), (x_2, y_2)$ are two

points where the transportation plan has support (i.e., $m(x, y) > 0$), then $x_2 > x_1$ implies $y_2 > y_1$. The interpretation is that optimal transportation paths of mass elements do not cross. For the case of matrix-valued distributions as in (3), this property may not hold in the same way. However, interestingly, a weaker monotonicity property holds for the supporting set of the optimal matrix transportation plan. The property is defined next and the precise statement is given in Proposition 2 below.

Definition 1—A set $\mathcal{S} \subset \mathbb{R}^2$ is called a ρ -monotonically nondecreasing, for $\rho > 0$, if for any two points $(x_1, y_1), (x_2, y_2) \in \mathcal{S}$, it holds that

$$(x_2 - x_1)(y_1 - y_2) \leq \rho.$$

A geometric interpretation for a ρ -monotonically non-decreasing set is that if $(x_1, y_1), (x_2, y_2) \in \mathcal{S}$ and $x_2 > x_1, y_1 > y_2$, then the area of the rectangle with vertices (x_i, y_j) ($i, j \in \{1, 2\}$) is not larger than ρ . The transportation plan of the scalar-valued optimal transportation problem with a quadratic cost has support on a 0-monotonically non-decreasing set.

Proposition 2—Given $\mu_0, \mu_1 \in \mathcal{F}$, let m be the optimal transportation plan in (10) with $c(x, y) = (x - y)^2$ and $\lambda > 0$. Then m has support on at most a $(4 \cdot \lambda)$ -monotonically nondecreasing set.

Proof: See the Appendix.

Further, the optimal transportation cost $\mathcal{T}_{2,\lambda}(\mu_0, \mu_1)$ satisfies:

1. $\mathcal{T}_{2,\lambda}(\mu_0, \mu_1) = \mathcal{T}_{2,\lambda}(\mu_1, \mu_0)$,
2. $\mathcal{T}_{2,\lambda}(\mu_0, \mu_1) \geq 0$,
3. $\mathcal{T}_{2,\lambda}(\mu_0, \mu_1) = 0$ if and only if $\mu_0 = \mu_1$.

Thus, although $\mathcal{T}_{2,\lambda}(\mu_0, \mu_1)$ can be used to compare matrix-valued densities, it is not a metric and neither is $\mathcal{T}_{2,\lambda}^{1/2}$ since the triangular inequality does not hold in general. We will introduce a slightly different formulation of a transportation problem which does give rise to a metric.

A. Optimal Transport on a Subset

In this subsection, we restrict attention to a certain subset of transport plans $M(\mu_0, \mu_1)$ and show that the corresponding optimal transportation cost induces a metric. More specifically, let

$$M_0(\mu_0, \mu_1) := \{m \mid m(x, y) = (\mu_0(x) \otimes \mu_1(y)) a(x, y), m \in M\}.$$

For $m(x, y) \in M_0(\mu_0, \mu_1)$,

$$\begin{aligned}\underline{\text{tr}}_0(\mathbf{m}(x, y)) &:= \boldsymbol{\mu}_1(y) / \text{tr}(\boldsymbol{\mu}_1(y)) \\ \underline{\text{tr}}_1(\mathbf{m}(x, y)) &:= \boldsymbol{\mu}_0(x) / \text{tr}(\boldsymbol{\mu}_0(x)).\end{aligned}$$

Given $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$, the “orientation” of the mass of $\mathbf{m}(x, y)$ is fixed. Thus, in this case, the optimal transportation cost is

$$\tilde{\mathcal{T}}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) := \min_{\mathbf{m} \in M_0(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)} \int \text{tr} \left(\left(c + \lambda \|(\underline{\text{tr}}_0 - \underline{\text{tr}}_1)\mathbf{m}(x, y)\|_{\text{F}}^2 \right) \mathbf{m} \right) dx dy. \quad (12)$$

Proposition 3—For $\mathcal{T}_{2,\lambda}$ as in (12) with $\lambda > 0$ and $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}$,

$$d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) := \left(\tilde{\mathcal{T}}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) \right)^{\frac{1}{2}} \quad (13)$$

defines a metric on $\mathcal{F}$.

Proof: It is straightforward to see that

$$d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) = d_{2,\lambda}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0) \geq 0$$

and that $d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) = 0$ if and only if $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$. We now show that the triangle inequality holds as well. For $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathcal{F}$, let

$$\begin{aligned}\mathbf{m}_{01}(x, y) &= \frac{\boldsymbol{\mu}_0(x)}{\text{tr}(\boldsymbol{\mu}_0(x))} \otimes \frac{\boldsymbol{\mu}_1(y)}{\text{tr}(\boldsymbol{\mu}_1(y))} m_{01}(x, y) \\ \mathbf{m}_{12}(y, z) &= \frac{\boldsymbol{\mu}_1(y)}{\text{tr}(\boldsymbol{\mu}_1(y))} \otimes \frac{\boldsymbol{\mu}_2(z)}{\text{tr}(\boldsymbol{\mu}_2(z))} m_{12}(y, z)\end{aligned}$$

denote the optimal transportation plan for the pairs $(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ and $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, respectively, where m_{01} and m_{12} are two (scalar-valued) joint densities on $\mathbb{R}^2$ with marginals $\text{tr}(\boldsymbol{\mu}_0)$, $\text{tr}(\boldsymbol{\mu}_1)$ and $\text{tr}(\boldsymbol{\mu}_1)$, $\text{tr}(\boldsymbol{\mu}_2)$, respectively. Given $m_{01}(x, y)$ and $m_{12}(y, z)$ there is a joint density function $m(x, y, z)$ on $\mathbb{R}^3$ with m_{01} and m_{12} as the marginals on the corresponding subspaces [3, p. 208]. We set

$$\mathbf{m}(x, y, z) = \frac{\boldsymbol{\mu}_0(x)}{\text{tr}(\boldsymbol{\mu}_0(x))} \otimes \frac{\boldsymbol{\mu}_1(y)}{\text{tr}(\boldsymbol{\mu}_1(y))} \otimes \frac{\boldsymbol{\mu}_2(z)}{\text{tr}(\boldsymbol{\mu}_2(z))} m(x, y, z)$$

and note that it has $\mathbf{m}_{01}$ and $\mathbf{m}_{12}$ as matrix-valued marginal distributions. Now, let $\mathbf{m}_{02}(x, z) = (\boldsymbol{\mu}_0(x)/\text{tr}(\boldsymbol{\mu}_0(x)) \otimes (\boldsymbol{\mu}_2(z)/\text{tr}(\boldsymbol{\mu}_2(z)))) \mathbf{m}_{02}(x, z)$ be the marginal of $\mathbf{m}(x, y, z)$ when tracing out the y -component. This $\mathbf{m}_{02}(x, z)$ is a possible transportation plan between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_2$. Hence

$$\begin{aligned}
d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_2) &\leq \left(\int_{\mathbb{R}^2} \left((x-z)^2 + \lambda \left\| \frac{\boldsymbol{\mu}_0(x)}{\text{tr}\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_2(z)}{\text{tr}\boldsymbol{\mu}_2(z)} \right\|_F^2 \right) m_{02} dx dz \right)^{\frac{1}{2}} \\
&= \left(\int_{\mathbb{R}^3} \left((x-z)^2 + \lambda \left\| \frac{\boldsymbol{\mu}_0(x)}{\text{tr}\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_2(z)}{\text{tr}\boldsymbol{\mu}_2(z)} \right\|_F^2 \right) m dx dy dz \right)^{\frac{1}{2}} \\
&= \left(\int_{\mathbb{R}^3} \left((x-y+y-z)^2 + \lambda \left\| \frac{\boldsymbol{\mu}_0(x)}{\text{tr}\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_1(y)}{\text{tr}\boldsymbol{\mu}_1(y)} + \frac{\boldsymbol{\mu}_1(y)}{\text{tr}\boldsymbol{\mu}_1(y)} - \frac{\boldsymbol{\mu}_2(z)}{\text{tr}\boldsymbol{\mu}_2(z)} \right\|_F^2 \right) m dx dy dz \right)^{\frac{1}{2}} \\
&\leq \left(\int_{\mathbb{R}^2} \left((x-y)^2 + \lambda \left\| \frac{\boldsymbol{\mu}_0(x)}{\text{tr}\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_1(y)}{\text{tr}\boldsymbol{\mu}_1(y)} \right\|_F^2 \right) m_{01} dx dy \right)^{\frac{1}{2}} + \left(\int_{\mathbb{R}^2} \left((y-z)^2 + \lambda \left\| \frac{\boldsymbol{\mu}_1(y)}{\text{tr}\boldsymbol{\mu}_1(y)} - \frac{\boldsymbol{\mu}_2(z)}{\text{tr}\boldsymbol{\mu}_2(z)} \right\|_F^2 \right) m_{12} dy dz \right)^{\frac{1}{2}} \\
&= d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) + d_{2,\lambda}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)
\end{aligned}$$

where the last inequality is due to the metric property of L_2 .

If $\lambda = 0$, then $\tilde{\mathcal{T}}_{2,0}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ is exactly the OMT cost between the scalar-valued densities $\text{tr}(\boldsymbol{\mu}_0)$ and $\text{tr}(\boldsymbol{\mu}_1)$ as was explained earlier. In particular, for an optimal transportation plan $m(x, y)$ between $\text{tr}(\boldsymbol{\mu}_0)$ and $\text{tr}(\boldsymbol{\mu}_1)$, the matrix-valued transportation plan $\tilde{\mathbf{m}}(x, y) = (\boldsymbol{\mu}_0(x)/\text{tr}(\boldsymbol{\mu}_0(x))) \otimes (\boldsymbol{\mu}_1(y)/\text{tr}(\boldsymbol{\mu}_1(y)))m(x, y)$ is optimal between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ which satisfies that $\tilde{\mathcal{T}}_{2,0}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) = \mathcal{T}_2(\text{tr}(\boldsymbol{\mu}_0), \text{tr}(\boldsymbol{\mu}_1))$. Thus, $\tilde{\mathcal{T}}_{2,0}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) = 0$ if and only if $\text{tr}(\boldsymbol{\mu}_0) = \text{tr}(\boldsymbol{\mu}_1)$. Hence, $\tilde{\mathcal{T}}_{2,0}$ fails to be a metric. Moreover, since for any $\lambda \geq 0$ it holds that $\tilde{\mathcal{T}}_{2,\lambda} \geq \tilde{\mathcal{T}}_{2,0}$, if $\text{tr}(\boldsymbol{\mu}_0) = \text{tr}(\boldsymbol{\mu}_1)$ then $\tilde{\mathcal{T}}_{2,0}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ also equals to zero.

Proposition 4—Given $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}$, let $\mathbf{m}$ be the optimal transportation plan in (13), then $\mathbf{m}$ has support on at most a $(2 \cdot \lambda)$ -monotonically non-decreasing set.

Proof: We need to prove that if $\mathbf{m}(x_1, y_1) = 0$ and $\mathbf{m}(x_2, y_2) = 0$, then $x_2 > x_1, y_1 > y_2$ implies

$$(y_1 - y_2)(x_2 - x_1) \leq 2\lambda. \quad (14)$$

Assume that $\mathbf{m}$ evaluated at the four points (x_i, y_j) with $i, j \in \{1, 2\}$, is as follows:

$$\mathbf{m}(x_i, y_i) = m_{ij} \cdot A_i \otimes B_j$$

with

$$A_i = \frac{\boldsymbol{\mu}_0(x_i)}{\text{tr}(\boldsymbol{\mu}_1(x_i))}, \quad B_i = \frac{\boldsymbol{\mu}_0(y_i)}{\text{tr}(\boldsymbol{\mu}_1(y_i))}$$

and $m_{11}, m_{22} > 0$. The steps of the proof are similar to those of Proposition 2 detailed in the Appendix: first, we assume that Proposition 4 fails and that

$$(y_1 - y_2)(x_2 - x_1) > 2\lambda.$$

Then we show that a smaller cost can be incurred by rearranging the “mass.” Consider the situation when $m_{22} > m_{11}$ first and let $\hat{\mathbf{m}}$ be a new transportation plan with

$$\begin{aligned}\hat{\mathbf{m}}(x_1, y_1) &= 0 \\ \hat{\mathbf{m}}(x_1, y_2) &= (m_{11} + m_{12}) \cdot A_1 \otimes B_2 \\ \hat{\mathbf{m}}(x_2, y_1) &= (m_{11} + m_{21}) \cdot A_2 \otimes B_1 \\ \hat{\mathbf{m}}(x_2, y_2) &= (m_{22} - m_{11}) \cdot A_2 \otimes B_2.\end{aligned}$$

Then, $\mathbf{m}, \hat{\mathbf{m}}$ have the same marginals at the four points, the cost incurred by $\mathbf{m}$ is

$$\sum_{i=1}^2 \sum_{j=1}^2 m_{ij} \left((x_i - y_j)^2 + \lambda \|A_i - B_j\|_{\text{F}}^2 \right) \quad (15)$$

and the cost incurred by $\hat{\mathbf{m}}$ is

$$\begin{aligned}& (m_{11} + m_{12}) \left((x_1 - y_2)^2 + \lambda \|A_1 - B_2\|_{\text{F}}^2 \right) \\ & + (m_{11} + m_{21}) \left((x_2 - y_1)^2 + \lambda \|A_2 - B_1\|_{\text{F}}^2 \right) \\ & + (m_{22} - m_{11}) \left((x_2 - y_2)^2 + \lambda \|A_2 - B_2\|_{\text{F}}^2 \right).\end{aligned} \quad (16)$$

To show that (15) is larger than (16), after canceling common terms, it suffices to show that

$$(y_1 - x_1)^2 + (y_2 - x_2)^2 + \lambda \|A_1 - B_1\|_{\text{F}}^2 + \lambda \|A_2 - B_2\|_{\text{F}}^2 \geq (y_2 - x_1)^2 + (y_1 - x_2)^2 + \lambda \|A_1 - B_2\|_{\text{F}}^2 + \lambda \|A_2 - B_1\|_{\text{F}}^2.$$

However, the above holds true since

$$\begin{aligned}& (y_1 - x_1)^2 + (y_2 - x_2)^2 + \lambda \|A_1 - B_1\|_{\text{F}}^2 + \lambda \|A_2 - B_2\|_{\text{F}}^2 \geq (y_1 - x_1)^2 + (y_2 - x_2)^2 \\ & = (y_1 - x_2)^2 + (y_2 - x_1)^2 + 2(x_2 - x_1)(y_1 - y_2) \\ & > (y_1 - x_2)^2 + (y_2 - x_1)^2 + 4\lambda \\ & \geq (y_1 - x_2)^2 + (y_1 - x_2)^2 + \lambda \left(\|A_1 - B_2\|_{\text{F}}^2 + \|A_2 - B_1\|_{\text{F}}^2 \right).\end{aligned}$$

The last inequality follows from:

$$\|A_1 - B_2\|_{\text{F}}^2 = \text{tr} \left(A_1^2 + B_2^2 - 2A_1 B_2 \right) \leq \text{tr} \left(A_1^2 + B_2^2 \right) \leq 2.$$

The case $m_{11} > m_{22}$ proceeds similarly.

V. Examples

We give two different examples where matrix-valued OMT can be directly applied. Both relate to spectral analysis of multivariable time series.³

A. Spectral Morphing

We first highlight the relevance of matrix-valued OMT to spectral analysis with a numerical example on spectral morphing. The idea is to model slowly time-varying changes in the spectral domain by geodesics in a suitable geometry (see e.g., [7], [8]). The use of geodesic interpolation can be thought of as a regularization technique. Indeed, geodesics smoothly shift spectral power across frequencies lessening the possibility of a fade-in fade-out artifacts and OMT, for scalar power spectra, has been used to this end in [7], [8]. Below we exemplify how geodesics appear in matrix-valued OMT.

Starting with $\mu_0, \mu_1 \in \mathcal{F}$ we approximate the geodesic between the two by constructing $N - 1$ points intermediate matrix densities. To this end, we set $\mu_{\tau_0} = \mu_0$ and $\mu_{\tau_N} = \mu_1$, and determine $\mu_{\tau_k} \in \mathcal{F}$ for $k = 1, \dots, N - 1$ as the solution to

$$\min_{\mu_{\tau_k}, 0 < k < N} \sum_{k=0}^{N-1} \mathcal{J}_{2,\lambda}(\mu_{\tau_{k+1}}, \mu_{\tau_k}). \quad (17)$$

As noted in Section III-D, this can be obtained numerically via convex programming. The present example uses

$$\mu_0 = \begin{bmatrix} 1 & 0 \\ 0.2e^{-j\theta} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{|a_0(e^{j\theta})|^2} & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} 1 & 0.2e^{j\theta} \\ 0 & 1 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.01 & 0 \\ 0 & \frac{1}{|a_1(e^{j\theta})|^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0.2 & 1 \end{bmatrix}$$

with

$$a_0(z) = (z^2 - 1.8\cos(\frac{\pi}{4})z + 0.9^2)(z^2 - 1.4\cos(\frac{\pi}{3})z + 0.7^2)$$

$$a_1(z) = (z^2 - 1.8\cos(\frac{\pi}{6})z + 0.9^2)(z^2 - 1.5\cos(\frac{2\pi}{15})z + 0.75^2)$$

shown in Fig. 1. Since the value of a power spectral density at each point in frequency is a 2×2 Hermitian matrix, we have used the (1, 1), (1, 2), and (2, 2) subplots to display the magnitude of the corresponding entries, i.e., $|\mu(1, 1)|$, $|\mu(1, 2)|$, ($= |\mu(2, 1)|$), and $|\mu(2, 2)|$, respectively, and the (2,1) subplot to display the phase $\angle\mu(1, 2)$ ($= -\angle\mu(2, 1)$).

The 3-D plots in Fig. 2 refer to (17), with $\lambda = 0.1$, for an approximation of a geodesic. The two boundary plots represent the power spectra μ_0 and μ_1 shown in blue and red, respectively, using the same convention about magnitudes and phases explained above. There are in total 7 power spectra μ_{τ_k} , $k = 1, \dots, 7$ shown along the geodesic between μ_0 and μ_1 , and the time-indices correspond to $\tau_k = k/8$. It is interesting to observe the smooth shift of the energy over the geodesic path from the one “channel” to the other while, at the same

³Matlab code is available at <http://www.ece.umn.edu/users/ningx015/research.html>.

time, the corresponding peak shifts from one frequency to another. One should bear in mind that the so-constructed geodesic is a non-parametric path interpolating/linking the given spectra.

B. Regularization Using Geodesics

Consider two time series

$$\begin{aligned} x_1(t) &= a_1(t) \cos(\theta_1(t)t + \phi_{1a}) + a_2(t) \cos(\theta_2(t)t + \phi_{1b}) + w_1(t) \\ x_2(t) &= a_2(t) \cos(\theta_1(t)t + \phi_{2a}) + a_1(t) \cos(\theta_2(t)t + \phi_{2b}) + w_2(t) \end{aligned}$$

for $t = 1, \dots, 2000$, both consisting of sinusoidal signals with time-varying amplitude and frequency (chirp-like) with added white noise $w_1(t)$ and $w_2(t)$. The amplitude $a_1(t)$ decreases from 1.2 to 0.1 while $a_2(t)$ increases from 0.1 to 1.2. Frequency $\theta_1(t)$ decreases from $(\pi/4)$ to $(\pi/4) - (\pi/30)$ while $\theta_2(t)$ increases from $(\pi/3)$ to $(\pi/3) + (\pi/30)$. Then $[w_1(t), w_2(t)]'$ is white, with independent components, and sampled from a zero-mean Gaussian distribution with

covariance $\begin{bmatrix} 3 & 1.5 \\ 1.5 & 3 \end{bmatrix}$. The initial phases of the sinusoids are randomly selected in $[0, 2\pi]$.

Since we are dealing with non-stationary time series, we truncate the observed time series, to segments of length equal to 200, to retain resolution. Thus, we let $x_{i,k}(t) := x_i(200k+t)$ with $i = 1, 2, k = 0, \dots, 9$ and process separately the segments $\{x_{i,k}(1), x_{i,k}(2), \dots, x_{i,k}(200)\}$. We obtain matrix-valued sample covariances $\{R_{k,-20}, \dots, R_{k,0}, \dots, R_{k,20}\}$ for each. We then determine autoregressive models based on these sample covariances and, thereby, the corresponding power spectral density functions. More specifically, for $\ell = 0, 1, \dots, 20$, we compute

$$R_{k,\ell} := \frac{1}{200} \sum_{i=\ell}^{200} \begin{bmatrix} x_{1,k}(i) \\ x_{2,k}(i) \end{bmatrix} [x_{1,k}(i-\ell), x_{2,k}(i-\ell)]$$

and let $R_{k,-\ell} = R_{k,\ell}'$. We then solve the Yule-Walker equations

$$R_{k,\ell} = \sum_{i=1}^{20} A_{k,i} R_{k,\ell-i}, \quad \text{for } \ell = 1, \dots, 20$$

for the autoregressive (matricial) coefficients $A_1, \dots, A_{20}$. We let $\Omega := R_0 - \sum_{i=1}^{20} A_{k,i} R_{k,-i}$ be the corresponding innovation variance. The estimated power spectral density function for the k th segment, denoted as $\hat{\mu}_{\tau_k}$, is given as $\hat{\mu}_{\tau_k}(\theta) = A_k(e^{j\theta})^{-1} \Omega (A_k(e^{j\theta})^*)^{-1}$ with

$$A_k(e^{j\theta}) = (I - A_{k,1}z - \dots - A_{k,20}z^{20})|_{z=e^{j\theta}}.$$

We scale $\hat{\boldsymbol{\mu}}_{\tau_k}$ so that the integral of its trace is normalized to one. Thus, the observation record is used to obtain 10 PSD's denoted as $\hat{\boldsymbol{\mu}}_{\tau_k}$, for $k = 1, \dots, 10$. These represent estimates of the spectral power at intermediary points in time. We change the time scale so that $\tau_1 = 0$ and $\tau_{10} = 1$. The spectrogram is shown in Fig. 3(a).

We construct an OMT-geodesic to regularize the estimated PSD's. This idea was proposed and carried out in [8] for scalar time series and scalar PSD's. For the present matrix-valued setting, the geodesic is obtained by solving

$$\min_{\boldsymbol{\mu}_{\tau_k}} \left\{ \sum_{k=1}^{10} \mathcal{F}_{2,\lambda}(\boldsymbol{\mu}_{\tau_k}, \hat{\boldsymbol{\mu}}_{\tau_k}) \mid \boldsymbol{\mu}_{\tau_k} \text{ are on an OMT geodesic} \right\}. \quad (18)$$

An explicit formula of the OMT geodesic is not available. However, in light of Proposition 2, (18) can be approximated for small λ as follows. Let $\hat{\boldsymbol{\mu}}_{\tau_k} = \text{tr}(\hat{\boldsymbol{\mu}}_{\tau_k})$ for $k = 1, \dots, 10$. These are scalar-valued PSD's. Let $\hat{M}_{\tau_k}$ denote the corresponding cumulative distribution functions. For λ small, following [8], we compute $\mu_0 := \text{tr}(\boldsymbol{\mu}_0)$ and $\mu_1 := \text{tr}(\boldsymbol{\mu}_1)$ via solving:

$$\min_{\mu_0, \mu_1} \sum_{k=1}^{10} \int_0^1 \left((1-\tau_k)M_0^{-1}(v) + \tau_k M_1^{-1}(v) - \hat{M}_{\tau_k}^{-1}(v) \right)^2 dv$$

with M_0 and M_1 representing the cumulative distribution function of μ_0 and μ_1 , respectively. Then, as was shown in [8], the μ_{τ_k} 's for $1 < k < 10$ can be computed via

$$M_{\tau_k}^{-1}(v) = (1-\tau_k)M_0^{-1}(v) + \tau_k M_1^{-1}(v).$$

The matrix-valued PSD's $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ are obtained by solving

$$\min_{\boldsymbol{\mu}_0, \boldsymbol{\mu}_1} \sum_{k=1}^{10} \int_0^1 \left\| (1-\tau_k) \frac{\boldsymbol{\mu}_0(M_0^{-1}(v))}{\mu_0(M_0^{-1}(v))} + \tau_k \frac{\boldsymbol{\mu}_1(M_1^{-1}(v))}{\mu_1(M_1^{-1}(v))} - \frac{\hat{\boldsymbol{\mu}}_{\tau_k}(\hat{M}_{\tau_k}^{-1}(v))}{\hat{\mu}_{\tau_k}(\hat{M}_{\tau_k}^{-1}(v))} \right\|_F^2 dv$$

and the $\boldsymbol{\mu}_{\tau_k}$'s for $1 < k < 10$ are computed via

$$\frac{\boldsymbol{\mu}_{\tau_k}(M_{\tau_k}^{-1}(v))}{\mu_{\tau_k}(M_{\tau_k}^{-1}(v))} = (1-\tau_k) \frac{\boldsymbol{\mu}_0(M_0^{-1}(v))}{\mu_0(M_0^{-1}(v))} + \tau_k \frac{\boldsymbol{\mu}_1(M_1^{-1}(v))}{\mu_1(M_1^{-1}(v))}.$$

We display this geodesic-fitted spectrogram in Fig. 3(b). It can be seen that the shift of energy from one channel to another and between resonant frequencies is smoother than that shown in Fig. 3(a) (which is a spectrogram based on matricial auto-regressive models).

In order to compare the resolution between the two techniques (spectrogram based on AR-modeling vs. geodesic regularization), we identify the frequency and directionality of peak power for the two power spectral densities and compare principal direction. This we explain next as the result is quite revealing and suggesting.

For each $\hat{\mu}_{\tau_k}(\theta)$, we find two frequencies θ_1 and θ_2 where the power spectral densities (PSD's) have locally maximal power, i.e., the two frequencies where $\text{tr}(\hat{\mu}_{\tau_k}(\theta))$ has the largest peaks. Then we compute the (normalized) eigenvectors corresponding to the dominant eigenvalues of $\mu_{\tau_k}(\theta_1)$ and $\mu_{\tau_k}(\theta_2)$, respectively. These eigenvectors are shown in Fig. 4(a) using black dashed lines. The red and green plots in Fig. 4(a) represent the path of the two eigenvectors as τ_k increases from 0 to 1. The axes in Fig. 4(a) correspond to the two channels/components of the time series and τ_k . (Should all the power be present in one of the two channels, the eigenvector would line up, accordingly, to one of the axes.) The values of the eigenvector when projected onto the two channels/axes, reflect the energy of the signals in the corresponding channels. Thus, in antenna-array applications, the direction of eigenvectors corresponds to the direction of a scatterer relative to the array. Statistical errors are reflected in the jagged nature of the paths when these are based on a spectrogram as in Fig. 4(a). However, when comparing with the eigenvectors of the OMT regularized spectrogram/AR-models μ_{τ_k} , the corresponding paths shown in Fig. 4(b) are smooth. Direct comparison between Fig. 4(a) and (b) highlights the potential advantages of using geodesics as a means to regularize power distribution in non-stationary time series.

VI. Conclusion

The geometry of Monge–Kantorovich optimal mass transport provides scalar densities with a natural metric structure (see [11], [12]; also [13] for a systems viewpoint and connections to image analysis and power spectra). Our interest has been in extending such a geometric structure to matrix-valued densities. To this end, we formulated one possible matrix-valued version of the Monge–Kantorovich transportation problem. Computations require convex programming and the framework directly extends the scalar case. An alternative generalization of the Monge–Kantorovich theory to a “non-commutative” setting has been given in the context of the theory of free-probabilities [14]. However, this may not be suitable for matrix-valued power distributions as it is not weak* continuous. Alternative non-commutative generalizations of the Wasserstein metric are given in [15]–[17]. Possible connections between the formulation herein and these alternative viewpoints is the subject of current investigation.

References

1. Monge, G. Open Library. De l’Imprimerie Royale; 1781. Mémoire sur la théorie des déblais et des remblais.
2. Kantorovich L. On the transfer of masses. Dokl Akad Nauk SSSR. 1942; 37:227–229.
3. Villani, C. Topics in optimal transportation. Vol. 58. Providence, RI: American Mathematical Society; 2003.
4. Ambrosio L. Lecture notes on optimal transport problems. Mathematical aspects of evolving interfaces. 2003:1–52.
5. Rachev, S.; Rüschendorf, L. Mass Transportation Problems: Theory. Vol. 1. New York: Springer-Verlag; 1998.

6. Ferrante A, Pavon M, Ramponi F. Hellinger versus Kullback–Leibler multivariable spectrum approximation. *IEEE Trans Autom Control*. Apr; 2008 53(4):954–967.
7. Jiang X, Ning L, Georgiou TT. Distances and Riemannian metrics for multivariate spectral densities. *IEEE Trans Autom Control*. Jul; 2012 57(7):1723–1735.
8. Jiang X, Luo Z, Georgiou T. Geometric methods for spectral analysis. *IEEE Trans Signal Process*. Mar; 2012 60(3):1064–1074.
9. Petz, D. *Quantum Information Theory and Quantum Statistics (Theoretical and Mathematical Physics)*. Berlin, Germany: Springer; 2008.
10. Boyd, S.; Vandenberghe, L. *Convex optimization*. Cambridge, MA: Cambridge Univ. Press; 2004.
11. Benamou J, Brenier Y. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*. 2000; 84(3):375–393.
12. Jordan R, Kinderlehrer D, Otto F. The variational formulation of the Fokker–Planck equation. *SIAM J Mathemat Anal*. 1998; 29(1):1–17.
13. Tannenbaum, E.; Georgiou, T.; Tannenbaum, A. Signals and control aspects of optimal mass transport and the Boltzmann entropy. *Proc. 49th IEEE Conf. Decision and Control*; 2010. p. 1885–1890.
14. Biane P, Voiculescu D. A free probability analogue of the Wasserstein metric on the trace-state space. *Geometric and Funct Anal*. 2001; 11(6):1125–1138.
15. Rieffel M. Metrics on state spaces. *Doc Math J DMV*. 1999; 4:559–600.
16. Andrea FD, Martinetti P. A view on optimal transport from noncommutative geometry. *SIGMA*. 2010; 6(057):24.
17. Martinetti P. Towards a Monge–Kantorovich metric in noncommutative geometry. *Zap Nauch Semin POMI*. 2013:411. arXiv preprint arXiv:1210.6573.

Biographies

Lipeng Ning received the B.S. and M.S. degrees in control science and engineering from Beijing Institute of Technology, Beijing, China, in 2006 and 2008, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2013.

He is currently a Postdoc Research Fellow in Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA. His research interests include system identification, spectral estimation, sparse signal recovery, and diffusion magnetic resonance imaging.

Tryphon T. Georgiou (F'00) received the Diploma in mechanical and electrical engineering from the National Technical University of Athens, Athens, Greece, in 1979 and the Ph.D. degree from the University of Florida, Gainesville, in 1983.

He is a faculty member in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, and the Vincentine Hermes-Luh Chair. He has served as a Co-Director of the Control Science and Dynamical Systems Center at the University of Minnesota (1990–2014), and on the Board of Governors of the Control Systems Society of the IEEE (2002–2005).

Dr. Georgiou has been a recipient of the George S. Axelby Outstanding Paper Award of the IEEE Control Systems Society for the years 1992, 1999, and 2003. He is a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).

Allen Tannenbaum (F'08) is a faculty member in computer science and applied mathematics at Stony Brook University, Stony Brook, NY, USA. He works in control, computer vision, and medical imaging.

Appendix

Appendix: Proof of Proposition 2

We need to show that if $m(x_1, y_1) = 0$ and $m(x_2, y_2) = 0$, then $x_2 > x_1, y_1 > y_2$ implies

$$(x_2 - x_1)(y_1 - y_2) \leq 4\lambda. \quad (19)$$

Without loss of generality, let

$$m(x_i, y_j) = m_{ij} \cdot A_{ij} \otimes B_{ij} \quad (20)$$

with $A_{ij}, B_{ij} \geq 0$, $\text{tr}(A_{ij}) = \text{tr}(B_{ij}) = 1$ and $i, j \in \{1, 2\}$. Note that m_{12} and m_{21} could be zero if $\mathbf{m}$ does not have support on the particular point. We assume that the condition in the proposition fails and that

$$(x_2 - x_1)(y_1 - y_2) > 4\lambda. \quad (21)$$

We then show that by rearranging the mass, the cost can be reduced.

We first consider the situation when $m_{22} = m_{11}$. By rearranging the value of $\mathbf{m}$ at the four points (x_i, y_j) with $i, j \in \{1, 2\}$, we construct a new transportation plan $\tilde{\mathbf{m}}$ at these four locations as follows:

$$\tilde{\mathbf{m}}(x_1, y_1) = 0 \quad (22a)$$

$$\tilde{\mathbf{m}}(x_1, y_2) = (m_{11} + m_{12}) \cdot \tilde{A}_{12} \otimes \tilde{B}_{12} \quad (22b)$$

$$\tilde{\mathbf{m}}(x_2, y_1) = (m_{11} + m_{21}) \cdot \tilde{A}_{21} \otimes \tilde{B}_{21} \quad (22c)$$

$$\tilde{\mathbf{m}}(x_2, y_2) = (m_{22} - m_{11}) \cdot A_{22} \otimes B_{22} \quad (22d)$$

where

$$\begin{aligned} \tilde{A}_{12} &= \frac{m_{11}A_{11} + m_{12}A_{12}}{m_{11} + m_{12}}, & \tilde{B}_{12} &= \frac{m_{11}B_{22} + m_{12}B_{12}}{m_{11} + m_{12}}, \\ \tilde{A}_{21} &= \frac{m_{11}A_{22} + m_{21}A_{21}}{m_{11} + m_{21}}, & \tilde{B}_{21} &= \frac{m_{11}B_{11} + m_{21}B_{21}}{m_{11} + m_{21}}. \end{aligned}$$

This new transportation plan $\tilde{\mathbf{m}}$ has the same marginals as $\mathbf{m}$ at x_1, x_2 and y_1, y_2 . The original cost incurred by $\mathbf{m}$ at these four locations is

$$\sum_{i=1}^2 \sum_{j=1}^2 m_{ij} \left((x_i - y_j)^2 + \lambda \|A_{ij} - B_{ij}\|_F^2 \right) \quad (23)$$

while the cost incurred by $\tilde{\mathbf{m}}$ is

$$\begin{aligned} & (m_{11} + m_{12}) \left((x_1 - y_2)^2 + \lambda \|\tilde{A}_{12} - \tilde{B}_{12}\|_F^2 \right) \\ & + (m_{11} + m_{21}) \left((x_2 - y_1)^2 + \lambda \|\tilde{A}_{21} - \tilde{B}_{21}\|_F^2 \right) \\ & + (m_{22} - m_{11}) \left((x_2 - y_2)^2 + \lambda \|A_{22} - B_{22}\|_F^2 \right). \end{aligned} \quad (24)$$

After simplification, to show that (23) is larger than (24), it suffices to show that

$$2m_{11}(x_2 - x_1)(y_1 - y_2) \quad (25)$$

is larger than

$$\lambda m_{11} \left(\sum_{i=1}^2 \sum_{j \neq i} \|\tilde{A}_{ij} - \tilde{B}_{ij}\|_F^2 - \sum_{i=1}^2 \|A_{ii} - B_{ii}\|_F^2 \right) \quad (26a)$$

$$+ \lambda m_{12} \left(\|\tilde{A}_{12} - \tilde{B}_{12}\|_F^2 - \|A_{12} - B_{12}\|_F^2 \right) \quad (26b)$$

$$+ \lambda m_{21} \left(\|\tilde{A}_{21} - \tilde{B}_{21}\|_F^2 - \|A_{21} - B_{21}\|_F^2 \right). \quad (26c)$$

From (21), it follows that the value in (25) is greater than $20\lambda m_{11}$. We derive upper bounds for each term in (26). First,

$$(26a) \leq \lambda m_{11} \left(\|\tilde{A}_{12} - \tilde{B}_{12}\|_F^2 + \|\tilde{A}_{21} - \tilde{B}_{21}\|_F^2 \right) \leq 4\lambda m_{11}$$

where the last inequality follows from the fact that:

$$\|A - B\|_F^2 = \text{tr}(A^2 - 2AB + B^2) \leq \text{tr}(A^2 + B^2) \leq 2$$

for any $A, B \succeq 0$ with $\text{tr}(A) = \text{tr}(B) = 1$. Now consider

$$\begin{aligned} \|\tilde{A}_{12} - \tilde{B}_{12}\|_F^2 - \|A_{12} - B_{12}\|_F^2 &= \text{tr} \left((\tilde{A}_{12} - \tilde{B}_{12} + A_{12} - B_{12})(\tilde{A}_{12} - \tilde{B}_{12} - A_{12} + B_{12}) \right) \\ &= \frac{m_{11}}{m_{11} + m_{12}} \left(\|A_{11} - B_{22}\|_F^2 - \|A_{12} - B_{12}\|_F^2 - \frac{m_{12}}{m_{11} + m_{12}} \|A_{11} - B_{22} - A_{12} + B_{12}\|_F^2 \right) \\ &\leq \frac{m_{11}}{m_{11} + m_{12}} \|A_{11} - B_{22}\|_F^2 \\ &\leq 2 \frac{m_{11}}{m_{11} + m_{12}} \end{aligned}$$

where the second equality follows from the definitions of $\tilde{A}_{12}$ and $\tilde{B}_{12}$ while the last inequality is obtained by bounding the terms in the trace. Thus, referring to expressions by the respective equation numbering

$$(26b) \leq 2\lambda m_{12} \frac{m_{11}}{m_{11} + m_{12}} \leq 2\lambda m_{11}.$$

In a similar manner, $(26c) \leq 2\lambda m_{11}$. Therefore

$$(26) \leq 8\lambda m_{11} < (25)$$

which implies that the cost incurred by $\tilde{m}$ is smaller than the cost incurred by m .

For the case where $m_{11} > m_{22}$, we can prove the claim by constructing a new transportation plan $\hat{m}$ with values

$$\begin{aligned}\hat{m}(x_1, y_1) &= (m_{11} - m_{22}) \cdot A_{11} \otimes B_{11} \\ \hat{m}(x_1, y_2) &= (m_{12} + m_{22}) \cdot \hat{A}_{12} \otimes \hat{B}_{12} \\ \hat{m}(x_2, y_1) &= (m_{21} + m_{22}) \cdot \hat{A}_{21} \otimes \hat{B}_{21} \\ \hat{m}(x_2, y_2) &= 0\end{aligned}$$

with

$$\begin{aligned}\hat{A}_{12} &= \frac{m_{12}A_{12} + m_{22}A_{11}}{m_{12} + m_{22}}, & \hat{B}_{12} &= \frac{m_{12}B_{12} + m_{22}B_{22}}{m_{12} + m_{22}} \\ \hat{A}_{21} &= \frac{m_{21}A_{21} + m_{22}A_{22}}{m_{21} + m_{22}}, & \hat{B}_{21} &= \frac{m_{21}B_{21} + m_{22}B_{11}}{m_{21} + m_{22}}.\end{aligned}$$

The rest of the proof is carried out in a similar manner.

Fig. 1.

Subplots (1,1), (1,2), and (2,2) show $\mu_i(1, 1)$, $|\mu_i(1, 2)|$ (same as $|\mu_i(2, 1)|$) and $\mu_i(2, 2)$.

Subplot (2,1) shows $\angle(\mu_i(2, 1))$ for $i \in \{0, 1\}$ in blue and red, respectively.

Fig. 2.

Interpolated results μ_{τ_k} for $k = 0, \dots, 8$ computed from (17) with μ_0 and μ_1 as the two boundary points: subplots (1, 1), (1, 2), and (2, 2) show $\mu_{\tau_k}(1, 1)$, $|\mu_{\tau_k}(1, 2)|$ (same as $|\mu_{\tau_k}(2, 1)|$) and $\mu_{\tau_k}(2, 2)$, subplot (2, 1) shows $\angle(\mu_{\tau_k}(2, 1))$.

Fig. 3.

(a) Shows the estimated spectrogram of the observed time series and (b) corresponds to the geodesic-fitted spectrogram.

Fig. 4.

In (a), the trajectories of the dominant eigenvector of $\hat{\mu}_{\tau_k}(\theta_1)$ and $\hat{\mu}_{\tau_k}(\theta_2)$ are shown in red and blue, respectively. The corresponding trajectories of $\mu_{\tau_k}(\theta_1)$ and $\mu_{\tau_k}(\theta_2)$ are shown in (b).