Title
On Matrix-Valued MongeKantorovich Optimal Mass Transport

Journal
IEEE Transactions on Automatic Control, 60(2)

ISSN
0018-9286

Authors
Ning, Lipeng
Georgiou, Tryphon T
Tannenbaum, Allen

Publication Date
2015-02-01

DOI
10.1109/tac.2014.2350171

Matrix-valued Monge-Kantorovich Optimal Mass Transport

Lipeng Ning, Tryphon T. Georgiou and Allen Tannenbaum

Abstract

We formulate an optimal transport problem for matrix-valued density functions. This is pertinent in the spectral analysis of multivariable time-series. The "mass" represents energy at various frequencies whereas, in addition to a usual transportation cost across frequencies, a cost of rotation is also taken into account. We show that it is natural to seek the transportation plan in the tensor product of the spaces for the two matrix-valued marginals. In contrast to the classical Monge-Kantorovich setting, the transportation plan is no longer supported on a thin zero-measure set.

I. Introduction

The formulation of optimal mass transport (OMT) goes back to the work of G. Monge in 1781 [1]. The modern formulation is due to Kantorovich in 1947 [2]. In recent years the subject is evolving rather rapidly due to the wide range of applications in economics, theoretical physics, probability, etc. Important recent monographs on the subject include [3], [4], [5].

Our interest in the subject of matrix-valued transport originates in the spectral analysis of multi-variable time-series. It is natural to consider the weak topology for power spectra. This is because statistics typically represent integrals of power spectra and hence a suitable form of continuity is desirable. Optimal mass transport and the geometry of the Wasserstein metric provide a natural framework for studying scalar densities. Thus, the scalar OMT theory was used in [6] for modeling slowly time-varying changes in the power spectra of time-series. The salient feature of matrix-valued densities is that power can shift across frequencies as well as across different channels via rotation of the corresponding eigenvectors. Thus, transport between matrix-valued densities requires that we take into account the cost of rotation as well as the cost of shifting power across frequencies.

Besides the formulation of a "non-commutative" Monge-Kantorovich transportation problem, the main results in the paper are that (1) the solution to our problem can be cast as a convex-optimization problem, (2) geodesics can be determined by convex programming, and (3) that 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.

II. Preliminaries on Optimal Mass Transport

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

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

The set $M(\mu_0,\mu_1)$ is not empty since $m(x,y)=\mu_0(x)\mu_1(y)$ is always a feasible solution. Probability densities can be thought of as distributions of mass and a cost $c(x,y)$ associated with transferring one unit of mass from one location x to y. For $c(x,y)=|x-y|^2$ the optimal transport 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

$$\mathcal{T}_2(\mu_0,\mu_1) := \inf_{m\in M(\mu_0,\mu_1)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)m(x,y)dxdy. \tag{1}$$

L. Ning and T.T. Georgiou are with the Department of Electrical & Computer Engineering, University of Minnesota, Minneapolis, MN 55455, {ningx015,tryphon}@umn.edu, A. Tannenbaum is with the Comprehensive Cancer Center and the Department of Electrical & Computer Engineering, University of Alabama, Birmingham, AL 35294, tannenba@uab.edu

Problem (1) is a linear programming problem with dual

$$\sup_{\phi,\psi}\left\{\int_{\mathbb{R}}\phi_0\mu_0-\phi_1\mu_1 dx \mid \phi_0(x)-\phi_1(y)\leq c(x,y)\right\} \tag{2}$$

see e.g., [3]. Moreover, for the quadratic cost function $c(x,y)=|x-y|^2$, $\mathcal{T}_2(\mu_0,\mu_1)$ can also be written explicitly in term of the cumulative distributions functions

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

as follows (see [3, page 75])

$$\mathcal{T}_2(\mu_0,\mu_1)=\int_0^1|F_0^{-1}(t)-F_1^{-1}(t)|^2 dt, \tag{3}$$

and the optimal joint probability density $m\in M(\mu_0,\mu_1)$ has support on $(x,T(x))$ where $T(x)$ is the sub-differential of a convex lower semi-continuous function. More specifically, $T(x)$ is uniquely defined by

$$F_0(x)=F_1(T(x)). \tag{4}$$

Finally, a geodesic $\mu_\tau(\tau\in[0,1])$ between μ_0 and μ_1 can be written explicitly in terms of the cumulative function F_τ defined by

$$F_\tau((1-\tau)x+\tau T(x))=F_0(x). \tag{5}$$

Then, clearly,

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

III. Matrix-valued Optimal Mass Transport

We consider the family

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

of Hermitian positive semi-definite, matrix-valued densities on $\mathbb{R}$, normalized so that their trace integrates to 1. We motivate a transportation cost to this matrix-valued setting and introduce a generalization of the Monge-Kantorovich OMT to matrix-valued densities.

A. Tensor product and partial trace

Consider two n-dimensional Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}_1$ with basis $\{u_1,\ldots,u_n\}$ and $\{v_1,\ldots,v_n\}$, respectively. 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. For $\boldsymbol{\rho}_0\in\mathcal{L}(\mathcal{H}_0)$ and $\boldsymbol{\rho}_1\in\mathcal{L}(\mathcal{H}_1)$, we denote their tensor product by $\boldsymbol{\rho}_0\otimes\boldsymbol{\rho}_1\in\mathcal{L}(\mathcal{H}_0\otimes\mathcal{H}_1)$. Formally, the latter is defined via

$$\boldsymbol{\rho}_0\otimes\boldsymbol{\rho}_1 \ :\ u\otimes v\mapsto \boldsymbol{\rho}_0 u\otimes\boldsymbol{\rho}_1 v.$$

Since our spaces are finite-dimensional this is precisely the Kronecker product of the corresponding matrix representation of the two operators.

Consider $\boldsymbol{\rho}\in\mathcal{L}(\mathcal{H}_0\otimes\mathcal{H}_1)$ which can be thought of as a matrix of size $n^2\times n^2$. The partial traces $\operatorname{tr}_{\mathcal{H}_0}$ and $\operatorname{tr}_{\mathcal{H}_1}$, or tr_0 and tr_1 for brevity, are linear maps

$$\begin{aligned} \boldsymbol{\rho}\in\mathcal{L}(\mathcal{H}_0\otimes\mathcal{H}_1) &\mapsto \operatorname{tr}_1(\boldsymbol{\rho})\in\mathcal{L}(\mathcal{H}_0)\\ &\mapsto \operatorname{tr}_0(\boldsymbol{\rho})\in\mathcal{L}(\mathcal{H}_1) \end{aligned}$$

that are defined as follows. Partition $\boldsymbol{\rho}$ into $n\times n$ block-entries and denote by $\boldsymbol{\rho}_{k\ell}$ the (k,ℓ)-th block $(1\leq k,\ell\leq n)$. Then the partial trace, e.g.,

$$\boldsymbol{\rho}_0:=\operatorname{tr}_1(\boldsymbol{\rho})$$

is the $n \times n$ matrix with

$$[\boldsymbol{\rho}_0]_{k\ell} = \operatorname{tr}(\boldsymbol{\rho}_{k\ell}), \text{ for } 1 \leq k, \ell \leq n.$$

The partial trace

$$\boldsymbol{\rho}_1 := \operatorname{tr}_0(\boldsymbol{\rho})$$

is defined in a similar manner for a corresponding partition of $\boldsymbol{\rho}$, see e.g., [7]. More specifically, for $1 \leq i, j \leq n$, let $\boldsymbol{\rho}^{ij}$ be a sub-matrix of $\boldsymbol{\rho}$ of size $n \times n$ with the (k, ℓ)-th entry $[\boldsymbol{\rho}^{ij}]_{k\ell} = [\boldsymbol{\rho}_{k\ell}]_{ij}$. Then the (i, j)-th entry of $\boldsymbol{\rho}_1$ is

$$[\boldsymbol{\rho}_1]_{ij} = \operatorname{tr}(\boldsymbol{\rho}^{ij}).$$

Thus

$$\operatorname{tr}_1(\boldsymbol{\rho}_0 \otimes \boldsymbol{\rho}_1) = \operatorname{tr}(\boldsymbol{\rho}_1)\boldsymbol{\rho}_0 \text{ and } \operatorname{tr}_0(\boldsymbol{\rho}_0 \otimes \boldsymbol{\rho}_1) = \operatorname{tr}(\boldsymbol{\rho}_0)\boldsymbol{\rho}_1.$$

B. Joint density for matrix-valued distributions

A naive attempt to define a joint probability density given marginals $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}_n$ is to consider a matrix-valued density with support on $\mathbb{R} \times \mathbb{R}$ such that $\boldsymbol{m} \geq 0$ and

$$\int_{\mathbb{R}} \boldsymbol{m}(x, y) dy = \boldsymbol{\mu}_0(x), \quad \int_{\mathbb{R}} \boldsymbol{m}(x, y) dx = \boldsymbol{\mu}_1(y). \tag{6}$$

However, in contrast to the scalar case, this constraint is not always feasible. To see this consider

$$\begin{aligned}
\boldsymbol{\mu}_0(x) &= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} \delta(x - x_1) + \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \delta(x - x_2), \\
\boldsymbol{\mu}_1(x) &= \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix} \delta(x - x_1) + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \delta(x - x_2).
\end{aligned}$$

It is easy to show that (6) cannot be met.

A natural definition for joint densities $\boldsymbol{m}$ that can serve as a transportation plan may be defined as follows. For $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$$\boldsymbol{m}(x, y) \text{ is } n^2 \times n^2 \text{ positive semi-definite matrix,} \tag{7a}$$

and with

$$\boldsymbol{m}_0(x, y) := \operatorname{tr}_1(\boldsymbol{m}(x, y)), \boldsymbol{m}_1(x, y) := \operatorname{tr}_0(\boldsymbol{m}(x, y)), \tag{7b}$$

one has

$$\int_{\mathbb{R}} \boldsymbol{m}_0(x, y) dy = \boldsymbol{\mu}_0(x), \int_{\mathbb{R}} \boldsymbol{m}_1(x, y) dx = \boldsymbol{\mu}_1(y). \tag{7c}$$

Thus, we denote by

$$\boldsymbol{M}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) := \Big\{ \boldsymbol{m} \mid \text{(7a) -- (7c) are satisfied} \Big\}.$$

For this family, given marginals, there is always an admissible joint distribution as stated in the following proposition.

Proposition 1: For any $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}_n$, the set $\boldsymbol{M}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ is not empty.

Proof: Clearly, $\boldsymbol{m} := \boldsymbol{\mu}_0 \otimes \boldsymbol{\mu}_1 \in \boldsymbol{M}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$. ∎

We next motivate a natural form for the transportation cost. This is a functional on the joint density as in the scalar case. However, besides a penalty on "linear" transport we now take into account an "angular" penalty as well.

C. Transportation cost

We interpret $\operatorname{tr}(\boldsymbol{m}(x,y))$ as the amount of "mass" that is being transferred from x to y. Thus, for a scalar cost function $c(x,y)$ as before, one may simply consider

$$\min_{\boldsymbol{m}\in\boldsymbol{M}(\boldsymbol{\mu}_0,\boldsymbol{\mu}_1)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\operatorname{tr}(\boldsymbol{m}(x,y))dxdy. \tag{8}$$

However, if $\operatorname{tr}(\boldsymbol{\mu}_0(x)) = \operatorname{tr}(\boldsymbol{\mu}_1(x))\ \forall x\in\mathbb{R}$, then the optimal value of (8) is zero. Thus (8) fails to quantify mismatch in the matricial setting.

For simplicity, throughout, we only consider marginals $\boldsymbol{\mu}$, which pointwise satisfy $\operatorname{tr}(\boldsymbol{\mu})>0$. $\operatorname{tr}(\boldsymbol{\mu}(x))$ is a scalar-valued density representing mass at location x while $\frac{\boldsymbol{\mu}(x)}{\operatorname{tr}(\boldsymbol{\mu}(x))}$ has trace 1 and contains directional information. Likewise, for a joint density $\boldsymbol{m}(x,y)$, assuming $\boldsymbol{m}(x,y)\neq 0$, we consider

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

Since $\underline{\operatorname{tr}}_0(\boldsymbol{m}(x,y))$ and $\underline{\operatorname{tr}}_1(\boldsymbol{m}(x,y))$ are normalized to have unit trace, their difference captures the directional mismatch between the two partial traces. Thus take

$$\operatorname{tr}(\|(\underline{\operatorname{tr}}_0-\underline{\operatorname{tr}}_1)\boldsymbol{m}(x,y)\|_{\mathrm{F}}^2\boldsymbol{m}(x,y))$$

to quantify the rotational mismatch. The above motivates the following cost functional that includes both terms, rotational and linear:

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

where $\lambda>0$ can be used to weigh in the relative significance of the two terms.

D. Optimal transportation problem

In view of the above, we now arrive at the following formulation of a matrix-valued version of the OMT, namely the determination of

$$\boldsymbol{\mathcal{T}}_{2,\lambda}(\boldsymbol{\mu}_0,\boldsymbol{\mu}_1) := \min_{\boldsymbol{m}\in\boldsymbol{M}(\boldsymbol{\mu}_0,\boldsymbol{\mu}_1)} \int_{\mathbb{R}\times\mathbb{R}} \operatorname{tr}\left((c+\lambda\|(\underline{\operatorname{tr}}_0-\underline{\operatorname{tr}}_1)\boldsymbol{m}\|_{\mathrm{F}}^2)\boldsymbol{m}\right)dxdy. \tag{9}$$

Interestingly, (9) can be cast as a convex optimization problem. We explain this next.

Since, by definition,

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

we deduce that

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

Now let $m(x,y)=\operatorname{tr}(\boldsymbol{m}(x,y))$ and let $\boldsymbol{m}_0(x,y)$ and $\boldsymbol{m}_1(x,y)$ be as in (7). The expression for the optimal cost in (9) can be equivalently written as

$$\begin{aligned}\min_{\boldsymbol{m}_0,\boldsymbol{m}_1,m}\Big\{\int \left(c(x,y)m(x,y)+\lambda\frac{\|\boldsymbol{m}_0-\boldsymbol{m}_1\|_{\mathrm{F}}^2}{m}\right)dxdy \mid\ & \boldsymbol{m}_0(x,y),\ \boldsymbol{m}_1(x,y)\geq 0\\ & \operatorname{tr}(\boldsymbol{m}_0(x,y))=\operatorname{tr}(\boldsymbol{m}_1(x,y))=m(x,y)\\ & \int \boldsymbol{m}_0(x,y)dy=\boldsymbol{\mu}_0(x)\\ & \int \boldsymbol{m}_1(x,y)dx=\boldsymbol{\mu}_1(y)\Big\}.\end{aligned} \tag{10}$$

Since, for $x > 0$,

$$\frac{(y-z)^2}{x}$$

is convex in the arguments x, y, z, it readily follows that the integral in (10) is a convex functional. All constraints in (10) are also convex and therefore, so is the optimization problem.

IV. On the geometry of Optimal Mass Transport

A standard result in the (scalar) OMT theory is that the transportation plan is the sub-differential of a convex function. As a consequence the transportation plan has support only on a monotonically non-decreasing zero-measure set. This is no longer true for the optimal transportation plan for matrix-valued density functions and this we discuss next.

In optimal transport theory for scalar-valued distributions, 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, then $x_2 > x_1$ implies $y_2 \geq y_1$. The interpretation is that optimal transportation paths do not cross. For the case of matrix-valued distributions as in (4), 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 3 below.

Definition 2: A set $\mathcal{S} \subset \mathbb{R}^2$ is called a λ-*monotonically non-decreasing*, for $\lambda > 0$, if for any two points $(x_1, y_1), (x_2, y_2) \in \mathcal{S}$, it holds that

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

A geometric interpretation for a λ-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 λ. The transportation plan of the scalar-valued optimal transportation problem with a quadratic cost has support on a 0-monotonically non-decreasing set.

Proposition 3: Given $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}$, let $\boldsymbol{m}$ be the optimal transportation plan in (9) with $\lambda > 0$. Then $\boldsymbol{m}$ has support on at most a $(4 \cdot \lambda)$-monotonically non-decreasing set.

Proof: See the appendix. ■

Then the optimal transportation cost $\mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ satisfies the following properties:

1) $\mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$=$\mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0)$,
2) $\mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) \geq 0$,
3) $\mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) = 0$ if and only if $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$.

Thus, although $\mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ can be used to compare matrix-valued densities, it is not a metric and neither is $\mathcal{T}_{2,\lambda}^{\frac{1}{2}}$ 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 $\boldsymbol{M}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ and show that the corresponding optimal transportation cost induces a metric. More specifically, let

$$\boldsymbol{M}_0(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) := \left\{ \boldsymbol{m} \mid \ \boldsymbol{m}(x, y) = \boldsymbol{\mu}_0(x) \otimes \boldsymbol{\mu}_1(y) a(x, y), \ \boldsymbol{m} \in \boldsymbol{M} \right\}.$$

For $\boldsymbol{m}(x, y) \in \boldsymbol{M}_0(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$,

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

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

$$\tilde{\boldsymbol{\mathcal{T}}}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) := \min_{\boldsymbol{m} \in \boldsymbol{M}_0(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)} \int \operatorname{tr} \left((c + \lambda \| (\underline{\mathrm{tr}}_0 - \underline{\mathrm{tr}}_1) \boldsymbol{m}(x, y) \|_{\mathrm{F}}^2) \boldsymbol{m} \right) dx dy. \tag{11}$$

Proposition 4: For $\boldsymbol{\mathcal{T}}_{2,\lambda}$ as in (11) and $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}$,

$$d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) := \left(\tilde{\boldsymbol{\mathcal{T}}}_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)\right)^{\frac{1}{2}} \tag{12}$$

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

Proof: It is straightforward to prove 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 will show that the triangle inequality also holds. For $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathcal{F}_n$, let

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

be 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 $\mathrm{tr}(\boldsymbol{\mu}_0)$, $\mathrm{tr}(\boldsymbol{\mu}_1)$ and $\mathrm{tr}(\boldsymbol{\mu}_1)$, $\mathrm{tr}(\boldsymbol{\mu}_2)$, 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, page 208]. We denote

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

then it has $\boldsymbol{m}_{01}$ and $\boldsymbol{m}_{12}$ as the matrix-valued marginal distributions.

Now, let $\boldsymbol{m}_{02}(x,z) = \frac{\boldsymbol{\mu}_0(x)}{\mathrm{tr}\,\boldsymbol{\mu}_0(x)} \otimes \frac{\boldsymbol{\mu}_2(z)}{\mathrm{tr}\,\boldsymbol{\mu}_2(z)} m_{02}(x,z)$ be the marginal distribution of $\boldsymbol{m}(x,y,z)$ when tracing out the y-component. Then $\boldsymbol{m}_{02}(x,z)$ is a candidate transportation plan between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_2$. Thus

$$\begin{aligned}
d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_2) &\leq \left(\int_{\mathbb{R}^2} \left((x-z)^2 + \lambda \| \frac{\boldsymbol{\mu}_0(x)}{\mathrm{tr}\,\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_2(z)}{\mathrm{tr}\,\boldsymbol{\mu}_2(z)} \|_{\mathrm{F}}^2\right) m_{02}\,dxdz\right)^{\frac{1}{2}}\\
&= \left(\int_{\mathbb{R}^3} \left((x-z)^2 + \lambda \| \frac{\boldsymbol{\mu}_0(x)}{\mathrm{tr}\,\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_2(z)}{\mathrm{tr}\,\boldsymbol{\mu}_2(z)} \|_{\mathrm{F}}^2\right) m\,dxdydz\right)^{\frac{1}{2}}\\
&= \left(\int_{\mathbb{R}^3} \left((x-y+y-z)^2 + \lambda \| \frac{\boldsymbol{\mu}_0(x)}{\mathrm{tr}\,\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_1(y)}{\mathrm{tr}\,\boldsymbol{\mu}_1(y)} + \frac{\boldsymbol{\mu}_1(y)}{\mathrm{tr}\,\boldsymbol{\mu}_1(y)} - \frac{\boldsymbol{\mu}_2(z)}{\mathrm{tr}\,\boldsymbol{\mu}_2(z)} \|_{\mathrm{F}}^2\right) m\,dxdydz\right)^{\frac{1}{2}}\\
&\leq \left(\int_{\mathbb{R}^2} \left((x-y)^2 + \lambda \| \frac{\boldsymbol{\mu}_0(x)}{\mathrm{tr}\,\boldsymbol{\mu}_0(x)} - \frac{\boldsymbol{\mu}_1(y)}{\mathrm{tr}\,\boldsymbol{\mu}_1(y)} \|_{\mathrm{F}}^2\right) m_{01}\,dxdy\right)^{\frac{1}{2}} +\\
&\quad \left(\int_{\mathbb{R}^2} \left((y-z)^2 + \lambda \| \frac{\boldsymbol{\mu}_1(y)}{\mathrm{tr}\,\boldsymbol{\mu}_1(y)} - \frac{\boldsymbol{\mu}_2(z)}{\mathrm{tr}\,\boldsymbol{\mu}_2(z)} \|_{\mathrm{F}}^2\right) m_{12}\,dydz\right)^{\frac{1}{2}}\\
&= d_{2,\lambda}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1) + d_{2,\lambda}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)
\end{aligned}$$

where the last inequality is from the fact that L_2-norm defines a metric. ■

Proposition 5: Given $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}$, let $\boldsymbol{m}$ be the optimal transportation plan in (12), then $\boldsymbol{m}$ has support on at most a $(2\cdot\lambda)$-monotonically non-decreasing set.

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

$$(y_1 - y_2)(x_2 - x_1) \leq 2\lambda. \tag{13}$$

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

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

with

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

and $m_{11}, m_{22} > 0$. The steps of the proof are similar to those of Proposition 3: first, we assume that Proposition 5 fails and that

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

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

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

Then, $\hat{\boldsymbol{m}}$ has the same marginals as $\boldsymbol{m}$ at the four points and the cost incurred by $\boldsymbol{m}$ is

$$\sum_{i=1}^{2}\sum_{j=1}^{2} m_{ij}\left((x_i - y_j)^2 + \lambda\|A_i - B_j\|_{\mathrm{F}}^2\right) \tag{14}$$

while the cost incurred by $\hat{\boldsymbol{m}}$ is

$$\begin{aligned}
&(m_{11} + m_{12})\left((x_1 - y_2)^2 + \lambda\|A_1 - B_2\|_{\mathrm{F}}^2\right)\\
+&(m_{11} + m_{21})\left((x_2 - y_1)^2 + \lambda\|A_2 - B_1\|_{\mathrm{F}}^2\right)\\
+&(m_{22} - m_{11})\left((x_2 - y_2)^2 + \lambda\|A_2 - B_2\|_{\mathrm{F}}^2\right).
\end{aligned} \tag{15}$$

After canceling the common terms, to show that (14) is larger than (15), it suffices to show that

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

The above holds since

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

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

V. EXAMPLE

We highlight the relevance of the matrix-valued OMT to spectral analysis by presenting an numerical example of spectral morphing. The idea is to model slowly time-varying changes in the spectral domain by geodesics in a suitable geometry (see e.g., [6], [8]). The importance of OMT stems from the fact that it induces a weakly continuous metric. Thereby, geodesics smoothly shift spectral power across frequencies lessening the possibility of a fade-in fade-out phenomenon. The classical theory of OMT allows constructing such geodesics for scalar-valued distributions. The example below demonstrates that we can now have analogous construction of geodesics of matrix-valued power spectra as well.

Starting with $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1 \in \mathcal{F}$ we approximate the geodesic between them by identifying $N-1$ points between the two. More specifically, we set $\boldsymbol{\mu}_{\tau_0} = \boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_{\tau_N} = \boldsymbol{\mu}_1$, and determine $\boldsymbol{\mu}_{\tau_k} \in \mathcal{F}_n$ for $k = 1, \ldots, N-1$ by solving

$$\min_{\boldsymbol{\mu}_{\tau_k}, 0<k<N} \sum_{k=0}^{N-1} \mathcal{T}_{2,\lambda}(\boldsymbol{\mu}_{\tau_{k+1}}, \boldsymbol{\mu}_{\tau_k}). \tag{16}$$

As noted in Section III-D, numerically this can be solved via a convex programming problem. The numerical example is based on the following two matrix-valued power spectral densities

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 1 & 0 \\ 0.2e^{-\mathrm{j}\theta} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{|a_0(e^{\mathrm{j}\theta})|^2} & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} 1 & 0.2e^{\mathrm{j}\theta} \\ 0 & 1 \end{bmatrix}$$
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.01 & 0 \\ 0 & \frac{1}{|a_1(e^{\mathrm{j}\theta})|^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0.2 & 1 \end{bmatrix}$$

with

$$\begin{aligned} a_0(z) = \; & (z^2 - 1.8\cos(\frac{\pi}{4})z + 0.9^2) \\ & (z^2 - 1.4\cos(\frac{\pi}{3})z + 0.7^2) \\ a_1(z) = \; & (z^2 - 1.8\cos(\frac{\pi}{6})z + 0.9^2) \\ & (z^2 - 1.5\cos(\frac{2\pi}{15})z + 0.75^2), \end{aligned}$$

shown in Figure 1. The value of a power spectral density at each point in frequency is a 2×2 Hermitian matrix. Hence, the $(1,1)$, $(1,2)$, and $(2,2)$ subplots display the magnitude of the corresponding entries, i.e., $|\boldsymbol{\mu}(1,1)|$, $|\boldsymbol{\mu}(1,2)|$ $(= |\boldsymbol{\mu}(2,1)|)$ and $|\boldsymbol{\mu}(2,2)|$, respectively. The $(2,1)$ subplot displays the phase $\angle\boldsymbol{\mu}(1,2)$ $(= -\angle\boldsymbol{\mu}(2,1))$.

The three dimensional plots in Figure 2 show the solution of (16) with $\lambda = 0.1$ which is an approximation of a geodesic. The two boundary plots represent the power spectra $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ shown in blue and red, respectively, using the same convention about magnitudes and phases. There are in total 7 power spectra $\boldsymbol{\mu}_{\tau_k}$, $k = 1, \ldots, 7$ shown along the geodesic between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$, and the time indices corresponds to $\tau_k = \frac{k}{8}$. It is interesting to observe the smooth shift of the energy from one "channel" to the other one over the geodesic path while the peak shifts from one frequency to another.

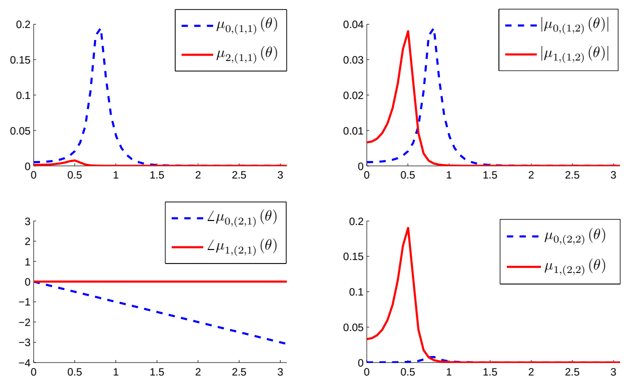

Fig. 1. Subplots (1,1), (1,2) and (2,2) show $\boldsymbol{\mu}_i(1,1)$, $|\boldsymbol{\mu}_i(1,2)|$ (same as $|\boldsymbol{\mu}_i(2,1)|$) and $\boldsymbol{\mu}_i(2,2)$. Subplot (2,1) shows $\angle(\boldsymbol{\mu}_i(2,1))$ for $i \in \{0,1\}$ in blue and red, respectively.

VI. Conclusions

This paper considers the optimal mass transportation problem of matrix-valued densities. This is motivated by the need for a suitable topology for the spectral analysis of multivariable time-series. It is well known that the OMT between scalar densities induces a Riemannian metric [9], [10] (see also [11] a systems viewpoint and connections to image analysis and metrics on power spectra). Our interest has been in extending such a Riemannian structure to matrix-valued densities. Thus, we formulate a "non-commutative" version of the Monge-Kantorovich transportation problem which can be cast as a convex-optimization problem. Interestingly, in contrast to the scalar case, the optimal transport plan is no longer supported on a set of measure zero. Versions of non-commutative Monge-Kantorovich transportation has been studied in the context of free-probability [12]. The relation of that to our formulation is still

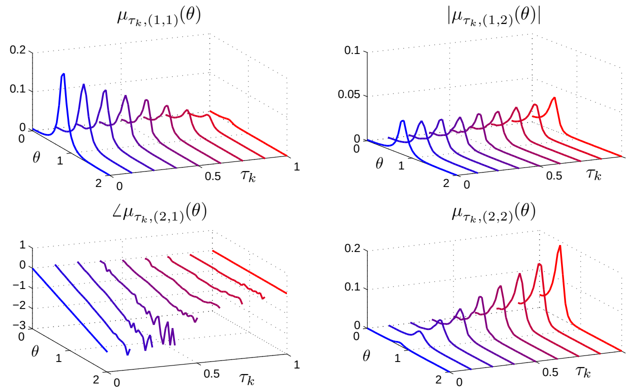

Fig. 2. The interpolated results $\boldsymbol{\mu}_{\tau_k}$ for $k = 0, \ldots, 8$ computed from (16) with $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ as the two boundary points: subplots (1,1), (1,2) and (2,2) show $\boldsymbol{\mu}_{\tau_k}(1,1), |\boldsymbol{\mu}_{\tau_k}(1,2)|$ (same as $|\boldsymbol{\mu}_{\tau_k}(2,1)|$) and $\boldsymbol{\mu}_{\tau_k}(2,2)$, subplot (2,1) shows $\angle(\boldsymbol{\mu}_{\tau_k}(2,1))$.

unclear. Finally, we note that if the matrix-valued distributions commute, then it is easy to check that our set-up reduces to that of a number of scalar problems, which is also the case in [12].

VII. Appendix: proof of Proposition 3

We need to prove that if $\boldsymbol{m}(x_1, y_1) \neq 0$ and $\boldsymbol{m}(x_2, y_2) \neq 0$, then $x_2 > x_1$, $y_1 > y_2$ implies

$$(x_2 - x_1)(y_1 - y_2) \leq 4\lambda. \tag{17}$$

Without loss of generality, let

$$\boldsymbol{m}(x_i, y_j) \quad = \quad m_{ij} \cdot A_{ij} \otimes B_{ij} \tag{18}$$

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

$$(x_2 - x_1)(y_1 - y_2) > 4\lambda, \tag{19}$$

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

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

$$\tilde{\boldsymbol{m}}(x_1, y_1) \quad = \quad 0 \tag{20a}$$
$$\tilde{\boldsymbol{m}}(x_1, y_2) \quad = \quad (m_{11} + m_{12}) \cdot \tilde{A}_{12} \otimes \tilde{B}_{12} \tag{20b}$$
$$\tilde{\boldsymbol{m}}(x_2, y_1) \quad = \quad (m_{11} + m_{21}) \cdot \tilde{A}_{21} \otimes \tilde{B}_{21} \tag{20c}$$
$$\tilde{\boldsymbol{m}}(x_2, y_2) \quad = \quad (m_{22} - m_{11}) \cdot A_{22} \otimes B_{22} \tag{20d}$$

where

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

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

$$\sum_{i=1}^{2} \sum_{j=1}^{2} m_{ij} \left((x_i - y_j)^2 + \lambda \|A_{ij} - B_{ij}\|_{\mathrm{F}}^2 \right) \tag{21}$$

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

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

After simplification, to show that (21) is larger than (22), it suffices to show that

$$2m_{11}(x_2-x_1)(y_1-y_2) \tag{23}$$

is larger than

$$\lambda m_{11}\left(\sum_{i=1}^{2}\sum_{j\neq i}\|\tilde{A}_{ij}-\tilde{B}_{ij}\|_{\mathrm{F}}^2-\sum_{i=1}^{2}\|A_{ii}-B_{ii}\|_{\mathrm{F}}^2\right) \tag{24a}$$

$$+\lambda m_{12}\left(\|\tilde{A}_{12}-\tilde{B}_{12}\|_{\mathrm{F}}^2-\|A_{12}-B_{12}\|_{\mathrm{F}}^2\right) \tag{24b}$$

$$+\lambda m_{21}\left(\|\tilde{A}_{21}-\tilde{B}_{21}\|_{\mathrm{F}}^2-\|A_{21}-B_{21}\|_{\mathrm{F}}^2\right). \tag{24c}$$

From the assumption in (19), the value of (23) $> 20\lambda m_{11}$. We derive upper bounds for each term in (24). First,

$$\text{(24a)} \leq \lambda m_{11}\left(\|\tilde{A}_{12}-\tilde{B}_{12}\|_{\mathrm{F}}^2+\|\tilde{A}_{21}-\tilde{B}_{21}\|_{\mathrm{F}}^2\right)\leq 4\lambda m_{11}$$

where the last inequality follows from the fact that for $A, B \geq 0$ and $\mathrm{tr}(A)=\mathrm{tr}(B)=1$,

$$\|A-B\|_{\mathrm{F}}^2=\mathrm{tr}(A^2-2AB+B^2)\leq \mathrm{tr}(A^2+B^2)\leq 2.$$

For an upper bound of (24b),

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

where the second equality follows from the definition of $\tilde{A}_{12}$ and $\tilde{B}_{12}$ while the last inequality is obtained by bounding the terms in the trace. Thus

$$\text{(24b)} \leq 2\lambda m_{12}\frac{m_{11}}{m_{11}+m_{12}}\leq 2\lambda m_{11}.$$

In a similar manner, (24c) $\leq 2\lambda m_{11}$. Therefore,

$$\text{(24)} \leq 8\lambda m_{11} < \text{(23)}$$

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

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

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

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.

Acknowledgments

This work was supported in part by grants from NSF, NIH, AFOSR, ONR, and MDA.

References

[1] G. Monge, *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.
[2] L. Kantorovich, "On the transfer of masses," in *Dokl. Akad. Nauk. SSSR*, vol. 37, 1942, pp. 227–229.
[3] C. Villani, *Topics in optimal transportation*. American Mathematical Society, 2003, vol. 58.
[4] L. Ambrosio, "Lecture notes on optimal transport problems," *Mathematical aspects of evolving interfaces*, pp. 1–52, 2003.
[5] S. Rachev and L. Rüschendorf, *Mass Transportation Problems: Theory*. Springer Verlag, 1998, vol. 1.
[6] X. Jiang, Z. Luo, and T. Georgiou, "Geometric methods for spectral analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1064–1074, 2012.
[7] D. Petz, *Quantum Information Theory and Quantum Statistics (Theoretical and Mathematical Physics)*. Berlin: Springer, 2008.
[8] X. Jiang, L. Ning, and T. T. Georgiou, "Distances and riemannian metrics for multivariate spectral densities," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1723–1735, 2012.
[9] J. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem," *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000.
[10] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker-Planck equation," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 1, pp. 1–17, 1998.
[11] E. Tannenbaum, T. Georgiou, and A. Tannenbaum, "Signals and control aspects of optimal mass transport and the Boltzmann entropy," in *49th IEEE Conference on Decision and Control*, 2010, pp. 1885–1890.
[12] P. Biane and D. Voiculescu, "A free probability analogue of the Wasserstein metric on the trace-state space," *Geometric and Functional Analysis*, vol. 11, no. 6, pp. 1125–1138, 2001.