Regularization and Interpolation of Positive Matrices

Yamamoto, Kaoru
Department of Automatic Control, Lund University

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

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

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

他

出版情報：IEEE Transactions on Automatic Control. 63 (4), pp.1208-1212, 2018-04. Institute of Electrical and Electronics Engineers (IEEE)
バージョン：
権利関係：

Regularization and Interpolation of Positive Matrices

Kaoru Yamamoto, Yongxin Chen, Lipeng Ning
Tryphon T. Georgiou, and Allen Tannenbaum

***Abstract*—We construct certain matricial analogues of mass transport for positive definite matrices of equal trace. The framework aims to devise ways of interpolating positive definite matrices that tradeoff between "aligning up their eigenstructure" and "scaling the corresponding eigenvalues". Motivation for the work is provided by power spectral analysis of multivariate time series where linear interpolation between matrix-valued power spectra generates push-pop unrealistic and undesirable artifacts.**

I. Introduction

The present paper is an attempt to develop a suitable matricial analogue of optimal mass transport (OMT). The basic problem of OMT refers to seeking a transportation plan that carries a given probability distribution to another so that a suitably defined transportation cost is minimized [1], [2], [3]. The original formulation of the problem by Monge in 1871 was motivated by civil engineering considerations, namely to transport dirt so as to level the ground. The mathematical problem achieved significant fame and notoriety due to inherent technical difficulties which persisted until the 1940's, at which time Kantorovich presented a relaxation of OMT in the form of a linear program. The relevance of this circle of ideas in the broader setting of resource allocation was already widely recognized and the impact secured a Nobel prize in Economics for Kantorovich in 1975. A new transformative phase of development in optimal mass transport began in the 1990's [4], [5], [6], [7] motivated by multitude of applications in physics, probability theory, image analysis, optimal control etc. In these, OMT was seen as an optimal control problem and stochastic formulations ensued (see [8], [9], [10] and the references therein).

The work presented herein is motivated by [11] where a matrix-valued formulation of OMT was introduced to address certain issues in spectral analysis of multivariable time series and system identification. More specifically, recall that the spectral content of scalar (slowly time-varying) time series is often displayed in the form of a spectrogram (time-frequency power distribution). The spectrogram illuminates any time-drift of the spectral content, but suffers from the perennial trade-off between variability and resolution (also known as the uncertainty principle of Fourier methods). Traditionally, windowing, kernel methods, and several other techniques have been used to enhance spectrograms. An alternative approach, highlighted in [12], is to interpolate spectral estimates in a suitable metric so as to track changes in the underlying spectral-content in a natural manner. Further, in [13] the authors have argued that OMT induces the appropriate topology (Wasserstein metric) where small perturbations correspond to small changes in estimated moments and vice versa. However, there has been no matrix-valued analogue of the Wasserstein metric. The work in [11] was perhaps the first attempt and had its basis in Kantorovich's idea of seeking a joint density (matrix-valued in our case) in a suitable product space, necessitating a prohibitively high computational burden. In the present paper we explore alternative ideas that are rooted in control. More specifically, we seek a dynamical evolution that allows rotation of eigenvectors as well as scales the corresponding eigenvalues so as to generate a path between end-point matrices. Accordingly, the choice of parameters (affecting cost) in an optimization problem promotes rotation or scaling.

Being inspired by the close connection between the heat equation, the Schrödinger equation and the scalar version of OMT, [2], [3], we formulate our problem based on some concepts in quantum mechanics. More specifically, probability density functions are exchanged for *density matrices* ρ, i.e., positive semidefinite Hermitian matrices with unit trace. Transport is then seen as a flow in the space of such matrices that minimizes a suitable cost functional.

The insight and techniques gained are aimed towards interpolating or regularizing sample covariances as well as matrix-valued power spectral densities of multivariate time series – they both reflect on how power varies with direction. More specifically, when dynamics that impact a vectorial process are slowly varying it is of interest to track changes by interpolating input-output short-window spectral estimates in a non-parametric manner, e.g., tracking frequencies of resonances in the spectral content along with a corresponding direction that couples specific entries. The ability to do so in a natural manner is even more enabling in array processing where, changes in spectral content of a matrix-valued spectrogram may be directly attributed to a moving scatterer. In particular, OMT-based interpolation prevents push pop artifacts as compared to linear interpolation (e.g., see [13] for scalar processes).

In the present work we focus on interpolation of matrices. We see this as the first step towards the development of more general transport between matrix-valued distributions (i.e., allowing for a spatial/frequency component). More specifically, at present, we focus on interpolating positive definite matrices in ways that allow controlling the correspondence between their eigenstructures. To this end, we decompose the tangent space of the cones of positive definite matrices into two subspaces, one that corresponds to rotating the eigenstructure and another that corresponds to scaling eigenvalues. In this way, by controlling these two complementing directions, we construct interpolating flows that have desired properties. Insights into setting up the problem are drawn from quantum mechanics where the principle object of study is the time-evolution of noncommutative operators (i.e., of matrices, when restricted to finite dimensions). Thus, in order to make the paper self-contained, we include a brief exposition of certain basic facts from quantum mechanics upon which we draw insights for the needed analysis.

Supported by AFOSR grants FA9550-12-1-0319 and FA9550-15-1-0045, NSF grant ECCS-1509387, NIH grants P41-RR-013218 and P41-EB-015902, and a fellowship through Memorial Sloan Kettering Cancer Center.

K. Yamamoto is with Department of Automatic Control, Lund University, Lund, Sweden. This author is a member of the LCCC Linnaeus Center and the ELLIIT Excellence Center at Lund University; email: k.yamamoto@ieee.org

Y. Chen is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY; email: chen2468@umn.edu

L. Ning is with Brigham and Women's Hospital (Harvard Medical School), MA.; email: lning@bwh.harvard.edu

T. T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA; email: tryphon@uci.edu

A. Tannenbaum is with the Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY; email: allen.tannenbaum@stonybrook.edu

The paper is structured as follows. In Section II, we present certain basic ideas of quantum mechanics that relate to material in the paper. In Section III we consider the tangent space of the cone of positive definite matrices. This leads to the material in Section IV on suitable cost functionals that promote a judicious balance between rotating the eigenstructure and scaling the eigenvalues. Numerical examples are discussed in Sections V and VI. Concluding remarks are provided in Section VII.

Notation: We denote by $\mathbb{H}$ the set of $n\times n$ Hermitian matrices, $\mathbb{S}$ the set of $n\times n$ skew-Hermitian matrices, and $\mathbb{H}_+$ the cones of positive-semidefinite matrices. Since matrices are $n\times n$ throughout the paper, we will not explicitly note dependence on n. The commutator of two square matrices A, B is denoted by $[A,B] := AB - BA$.

II. Quantum insights

The development draws on concepts from quantum mechanics and, therefore, we begin with a brief account of basics (see standard references, e.g., [15], for more).

A. Schrödinger equation

The evolution of closed quantum systems, i.e., one having no interaction with other quantum systems or a heat bath, is given by the *time-dependent Schrödinger equation* [15]:

$$\frac{\partial \psi}{\partial t} = X\psi \tag{1}$$

where $\psi \in \mathbb{C}^n$ and X is a skew Hermitian matrix[1]. Equation (1) describes a unitary evolution for the wave function, in our case, vector ψ; the quantum system is in a "pure state" in that the density matrix $\rho = \psi\psi^*$ has rank 1. A system is in a mixed state when the density matrix

$$\rho = \sum_k \lambda_k \psi_k \psi_k^*$$

with $\sum_k \lambda_k = 1$, has rank higher than 1. Either way, the density matrix evolves according to

$$\frac{\partial \rho}{\partial t} = X\rho - \rho X =: [X, \rho], \tag{2}$$

where the derivative is thought of entry-wise. It is evident that if the system is in a pure state, it remains so, as the rank of

$$\rho(t) = e^{Xt}\psi(0)\psi(0)^* e^{-Xt}$$

remains invariant. Likewise, if the system is in a mixed state

$$\begin{aligned}\rho(t) &= \sum_k \lambda_k(t)\psi_k(t)\psi_k^*(t)\\ &= e^{Xt}\left(\sum_k \lambda_k(t)\psi_k(0)\psi_k(0)^*\right)e^{-Xt},\end{aligned}$$

the eigenvalues $\lambda_k(t)$ of the density matrix remain invariant over time t, i.e., $\lambda_k(t) = \lambda_k(0)$ for all t. Thus, the evolution governed by (2) rotates in the same way the complete set of eigenvectors of the density matrix without changing the corresponding eigenvalues.

[1]More generally, ψ belongs to a Hilbert space and accordingly X is a skew Hermitian operator on that same Hilbert space. Typically X is expressed as $-\frac{i}{\hbar}H$ where H is a Hamiltonian (Hermitian) operator and $\hbar$ is the reduced Plank constant.

B. Evolution of density matrices

Decoherence and changes in the spectrum of ρ are typically modeled through coupling with an *ancilla* which is another quantum system. The state of the original system is then obtained by *tracing out* the ancillatory component of the joint density operator. *Lindblad's equation*, [16], [17], describes precisely such an evolution for the component of the original system. The Lindblad equation has the form

$$\frac{\partial \rho}{\partial t} = [X,\rho] - \sum_k \left(\frac{1}{2}(Y_k\rho + \rho Y_k) - Z_k\rho Z_k^*\right)$$

where $Y_k = Z_k^* Z_k$. The presence of $-Z\rho Z^*$ ensures that $\mathrm{trace}(\rho)$ remains constant while both the eigenvalues and the eigenstructure may change over time. Alternatively, one may consider more generally

$$\frac{\partial \rho}{\partial t} = [X,\rho] + u \tag{3}$$

where $\mathrm{trace}(u) = 0$ so as to preserve the trace of ρ. In fact, in what follows, we will do exactly that and consider flows in directions corresponding to traceless component u. Positivity of the flow will be ensured as an added (convex) condition and will not be intrinsically encoded in u (as in the Linblad equation where the right hand side is linear in ρ).

III. Trace-preserving linear flow on positive matrices

Consider the set of positive semidefinite matrices that are normalized to have trace one,

$$\mathcal{D} := \{\rho \in \mathbb{H}_+ \mid \mathrm{trace}(\rho) = 1\}.$$

As we noted earlier, we seek flows on $\mathcal{D}$ that preserve trace. The tangent space of $\mathcal{D}$ at any $\rho \in \mathcal{D}$ is

$$T_\rho = \{u \mid u \in \mathbb{H}, \mathrm{trace}(u) = 0\}.$$

In this, the subspace of traceless Hermitian components[2]

$$R_\rho := \{[X,\rho] \mid X \in \mathbb{S}\},$$

is responsible for rotating the eigensubspaces of ρ as we have noted in the previous section.

We now seek to identify the orthogonal complement of R_ρ so as to isolate the two directions that are responsible for rotation of eigenvectors and scaling of eigenvalues[3]. To this end, consider $u \in \mathbb{H}$ such that

$$\mathrm{trace}(u[X,\rho]) = 0 \quad \forall X \in \mathbb{S}. \tag{4}$$

Since the trace is invariant under cyclic permutations,

$$\mathrm{trace}([u,\rho]X) = 0 \quad \forall X \in \mathbb{S}. \tag{5}$$

But $[u,\rho]$ is already in $\mathbb{S}$, hence it is zero and therefore u must commute with ρ. Thus, from (4) we have that u is in the orthogonal complement of R_ρ. We summarize our conclusion as follows.

Proposition 1: The tangent space T_ρ of $\mathcal{D} := \{\rho \in \mathbb{H}_+ \mid \mathrm{trace}(\rho) = 1\}$ at $\rho \in \mathcal{D}$ can be decomposed as the direct sum

$$T_\rho = R_\rho \oplus C_\rho$$

of orthogonal components

$$\begin{aligned}&R_\rho = \{[X,\rho] \mid X \in \mathbb{S}\} \text{ and}\\ &C_\rho = \{u \mid u \in \mathbb{H},\ [u,\rho] = 0, \text{ and } \mathrm{trace}(u) = 0\}.\end{aligned}$$

[2]If $X \in \mathbb{S}$ and $\rho \in \mathbb{H}$, both $[X,\rho]^* = [X,\rho] \in \mathbb{H}$ and $\mathrm{trace}([X,\rho]) = 0$.
[3]The inner product of two matrices is $\langle u, v\rangle = \mathrm{trace}(u'v)$

If $X(t) \in \mathbb{S}$ and $u(t) \in C_{\rho(t)}$ for all t, then $\mathrm{trace}(\rho(t))$ remains constant with t.

IV. Interpolating flows between ρ_0 and ρ_1

Following on the previous rationale we may seek paths between density matrices ρ_0 and ρ_1, that minimize a suitable cost functional that allows trading off between the eigenstructure rotation specified by $X(t) \in \mathbb{S}$ and the eigenvalue scaling affected by $u(t) \in \mathbb{H}$. Assuming a constant rate of rotation and a constant "drift" of the spectrum, we formulate the following problem:

Problem 1: Minimize $\|X\|_2 + \epsilon\|Z\|_2$ subject to

$$\dot{\rho}(t) = [X, \rho(t)] + \underbrace{e^{Xt} Z e^{-Xt}}_{u(t)}, \text{ for } \rho(0) = \rho_0,\ \rho(1) = \rho_1,$$

along with $X \in \mathbb{S}$, $Z \in \mathbb{H}$, $[\rho_0, Z] = 0$, and $\mathrm{trace}(Z) = 0$.

In the above, the parameter $\epsilon \in [0, \infty)$ represents a weight that helps trading off the two alternative mechanisms for shifting eigenvalues and eigensubspaces to match the two end-point matrices. In general, "rotation" and "scaling" may vary over time. We readily verify that $u(t)$ in Problem 1 commutes with $\rho(t)$ as long as Z commutes with ρ_0. Thence, it can be seen that Problem 1 is equivalent to

$$\text{minimize} \quad \|X\|_2 + \epsilon\|e^{-X}\rho_1 e^{X} - \rho_0\|_2 \tag{6a}$$

$$\text{subject to} \quad X + X^T = 0, [\rho_0, e^{-X}\rho_1 e^{X} - \rho_0] = 0. \tag{6b}$$

This is a constrained nonlinear optimization problem that we approach, numerically, using `fmincon` in MATLAB.

V. Example: Interpolation of density matrices

In this section we highlight how interpolation is effected, via solving Problem 1, as a proof of concept. Starting from two end-point density matrices, the framework allows constructing alternative paths connecting the two where one may tradeoff the two possible ways that the transition from one to the other may take place, i.e., allowing for the eigenvalues to adjust by "scaling" and the eigenvectors to "rotate," respectively.[4]

Example 1: Consider the two density matrices

$$\rho_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \rho_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

On one end, a choice of ϵ (vanishingly small) in Problem 1 leads to a path that displays a fade-in/fade-out effect of scaling the eigenvalues, so as to connect the two end-points (Fig. 1a). No rotation of eigenvectors takes place. On the other end, for ϵ large (e.g., $\epsilon = 10$), we obtain a path where "rotation" of the eigenvectors is less costly (Fig. 1b); it is worth noting that in this case, since both matrices have rank one, *the path remains rank one*.

Motivation for our framework stems from multivariable time series analysis where power is often associated (e.g., in sensor arrays or radar) with the position of dominant scatterers. Fade-in-fade-out effects when interpolating or smoothing multivariable spectra are obviously undesirable as they create artifacts. Such fade-in-fade-out effects may be erroneously interpreted as due to the presence of additional scatterers beyond those that are present. The above rudimentary example may correspond to the case of two sensors reading a constant-frequency echo from a scatterer that changes its relative position with respect to the two. When recorded signals are correlated, the matrix-valued power spectrum at the corresponding frequency has (approximately) rank one. Likewise, movement of the scatterer that corresponds to a path between the two matrices, ought to have rank one (as in Fig. 1b). This exemplifies the need for paths that avoid push-pop for the corresponding eigenvalues (as linear interpolation would – one eigenvalue reducing while another increasing at the same time).

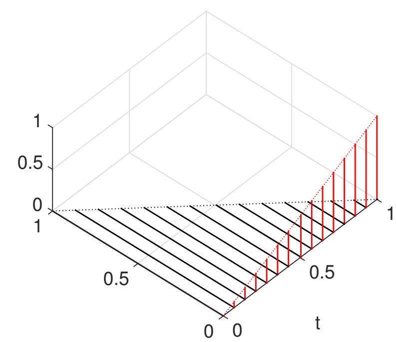
Example 2: In this we highlight a situation with matrices of higher dimension (3 in this case). Figure 2 shows a path between two matrices below

$$\rho_0 = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \text{ and } \rho_1 = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}.$$

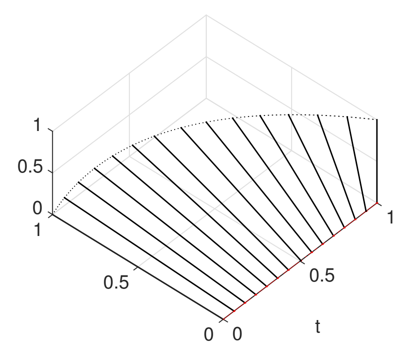
The solution to (6) with $\epsilon = 10$ is $\rho(t) = e^{Xt}(\rho_0 + Zt)e^{-Xt}$ with

$$X = \begin{bmatrix} 0 & 0 & 2.2 \\ 0 & 0 & 2.2 \\ -2.2 & -2.2 & 0 \end{bmatrix} \text{ and } Z = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & -0.3 \end{bmatrix}.$$

In this case, we see that both rotation and scaling take place simultaneously.



(a) Interpolation via adjusting eigenvalues.



(b) Interpolation via rotating eigenvectors.

Fig. 1: Solutions are obtained by solving Problem 1: for illustration, eigenvectors (shown) are scaled in proportion to corresponding eigenvalues. Fig. 1a shows a push-pop effect where eigenvalues are re-scaled whereas Fig. 1b indicates a correspondence through rotation of the eigenstructure through a path of rank-1 matrices.

[4]The examples we present involve matrices of sizes 2×2 and 3×3. This is solely because we cannot display in a suggestive manner results in higher dimensions. The computational burden in higher dimensions scales reasonably well (interpolating 20×20 matrices requires of the order of 200 [sec] using general purpose solvers on a laptop with Intel Core i7).

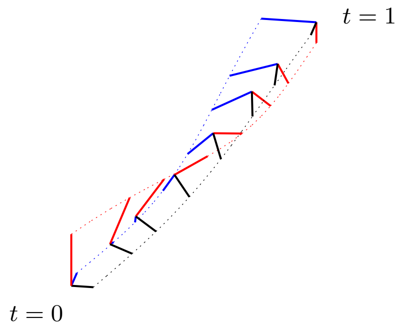


Fig. 2: Solutions are obtained by solving Problem 1: for illustration, eigenvectors (shown) are scaled in proportion to corresponding eigenvalues. It is seen how eigenstructure rotates while eigenvalues scale.

VI. Example: Regularization of noisy paths

Besides interpolation problems, i.e., finding a path for $\rho(t)$ connecting two density matrices ρ_0 and ρ_1, the approach allows solving regularization problems where a smooth path is constructed to smooth out noisy measurements. More specifically, given a noisy data set

$$\{\tilde{\rho}(t_i) \mid 0 \leq t_1 \leq \ldots \leq t_N \leq 1,\}$$

we seek a smooth path $\rho(t)$ that approximately fits the data in a suitable sense. The key is to parameterize the path in a way consistent with the two "orthogonal" actions of rotating eigenvectors and scaling eigenvalues (as both may be needed), and penalize one more (typically, scaling). To this end, we propose the following:

Problem 2: Minimize, over choice of ρ_0, X, Z,

$$\sum_{i=1}^{N} \left\| e^{Xt_i}(\rho_0 + Zt_i)e^{-Xt_i} - \tilde{\rho}(t_i) \right\|_2,$$

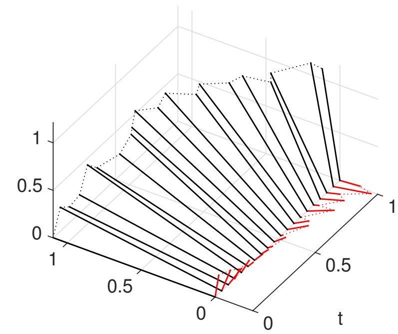
subject to $X \in \mathbb{S}$, $Z \in \mathbb{H}$, $\rho(\cdot) \geq 0$, $[\rho_0, Z] = 0$ and $\text{trace}(Z) = 0$.

The outcome is shown in Figure. 3. For illustration purposes the data set $\tilde{\rho}(t_i)$ is generated by adding a symmetric matrix-valued (uniform) noise $w(t)$ to a nominal flow $e^{Xt}\rho_0 e^{-Xt}$ for $t_i \in \{0.05, 0.1, 0.15, \ldots, 1\}$ where

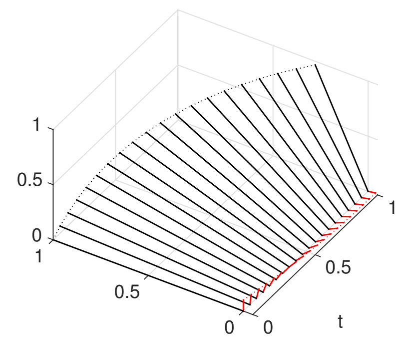
$$\rho_0 = \begin{bmatrix} 1.0 & 0 \\ 0 & 0.1 \end{bmatrix} \text{ and } X = \begin{bmatrix} 0 & -1.6 \\ 1.6 & 0 \end{bmatrix}.$$

VII. Concluding remarks

We propose an approach to constructing flows on density matrices. This is intended for interpolation and regularization of sample covariances and power spectra of multivariable time series. The general approach is control theoretic in that we select the flow (tangent direction) that minimizes a suitable cost functional. The choice of functional allows trading off the two basic mechanisms (rotating eigenvectors vs. scaling eigenvalues). Judicious balance between aligning up the eigenstructure and scaling the eigenvalues is necessitated by the fact that one of the two mechanisms alone may not suffice in generic situations. The choice of the parameter ϵ that dictates the respective tradeoff must be based on the application and priors on the problem. In closing, we refer the interested reader to [14] for a parallel alternative formulation of matricial OMT, which also draws on the connection with quantum mechanics, but along a different angle altogether.



(a) Data: noisy matrices $\tilde{\rho}(t_i)$.



(b) Regularized path $\rho(t_i)$.

Fig. 3: Regularization of noisy matricial data with eigenvectors scaled according to corresponding eigenvalues.

References

[1] S. Rachev and L. Rüschendorf, *Mass Transportation Problems*, Volumes I and II, Probability and Its Applications, Springer, New York, 1998.

[2] C. Villani, *Topics in Optimal Transportation*. American Mathematical Soc., 2003, no. 58.

[3] C. Villani, *Optimal Transport: Old and New*. Springer, 2008, vol. 338.

[4] R. J. McCann, "A convexity principle for interacting gases," *Advances in mathematics*, vol. 128, no. 1, pp. 153–179, 1997.

[5] W. Gangbo and R. J. McCann, "The geometry of optimal transportation," *Acta Mathematica*, vol. 177, no. 2, pp. 113–161, 1996.

[6] 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.

[7] J.-D. 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.

[8] Y. Chen, T.T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part I" *IEEE Transactions on Automatic Control* 61.5 (2016): 1158-1169.

[9] Y. Chen, T.T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part II," *IEEE Transactions on Automatic Control* 61.5 (2016): 1170-1180.

[10] Y. Chen, T.T. Georgiou, and M. Pavon, "Optimal transport over a linear dynamical system," arXiv preprint arXiv:1502.01265 (2015), *IEEE Transactions on Automatic Control*, to appear.

[11] L. Ning, T. T. Georgiou, and A. Tannenbaum, "On matrix-valued Monge–Kantorovich optimal mass transport," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 373–382, 2015.

[12] T.T. Georgiou, “The meaning of Distances in Spectral Analysis,” Plenary Lecture, IEEE Conf. on Decision and Control, New Orleans, December 2007, Control Systems Society Lecture Library http://www.ieeecss-oll.org (2007).

[13] T.T. Georgiou, J. Karlsson, and S. Mir. Takyar, “Metrics for power spectra: an axiomatic approach,” *IEEE Transactions on Signal Processing*, vol. 57(3): 859-67 (2009).

[14] Y. Chen, T. Georgiou, and A. Tannenbaum, “Matrix optimal mass transport: a quantum mechanical approach,” 2016, https://arxiv.org/abs/1610.03041.

[15] S. Gustafson and I. M. Sigal, *Mathematical Concepts of Quantum Mechanics*, Springer, New York, 2011.

[16] G. Lindblad, “On the generators of quantum dynamical semigroups,” *Communications in Mathematical Physics*, vol. 48, no. 2, pp. 119–130, 1976.

[17] P. Pearle, “Simple derivation of the Lindblad equation,” *European Journal of Physics*, vol. 33, no. 4, pp. 805, 2012.