Metrics for Power Spectra: An Axiomatic Approach

Tryphon T. Georgiou, *Fellow, IEEE*, Johan Karlsson, and Mir Shahrouz Takyar, *Student Member, IEEE*

***Abstract*—We present an axiomatic framework for seeking distances between power spectral density functions. The axioms require that the sought metric respects the effects of additive and multiplicative noise in reducing our ability to discriminate spectra, as well as they require continuity of statistical quantities with respect to perturbations measured in the metric. We then present a particular metric which abides by these requirements. The metric is based on the Monge-Kantorovich transportation problem and is contrasted with an earlier Riemannian metric based on the minimum-variance prediction geometry of the underlying time-series. It is also being compared with the more traditional Itakura-Saito distance measure, as well as the aforementioned prediction metric, on two representative examples.**

***Index Terms*—Geodesics, geometry of spectral measures, metrics, power spectra, spectral distances.**

I. Introduction

A key element of any quantitative scientific theory is a well defined and natural metric. A model for the development of such metrics is provided, in the context of information theory and statistics, in the work of Fisher, Rao, Amari, Centsov, and many others, via an axiomatic approach where the sought metric is identified on the basis of a natural set of axioms—the main one being the contractiveness of stochastic maps. The subject of the present paper is not the geometry of information, but instead, the possibility of analogous geometries for power spectra starting from a similar axiomatic rationale. Specifically, we seek a metric between power spectra which is contractive when noise is introduced, since intuitively, noise impedes our ability to discriminate. Further, we require that statistics are continuous with respect to spectral uncertainty quantified by the sought metric. We build on [25] where a variety of potential metrics were studied using complex analysis. The focus of the current paper is twofold, firstly to propose a natural set of axioms that geometries for power spectra must satisfy, and secondly to present a particular candidate which abides by the stated axioms. This metric is based on the Monge-Kantorovich transportation problem and represents a relaxation of Wasserstein distances so as to be applicable to power spectra.

In Section II, we outline and discuss the axiomatic framework. In Section III, we contrast the present setting with two alternatives, first the axiomatic basis of Information Geometry and then with a geometry that is inherited by linear prediction theory. In Section IV, we present basic facts of the Monge-Kantorovich transportation problem which are then utilized in Section V in order to develop a suitable family of metrics satisfying the axioms of the sought spectral geometry.

Manuscript received December 08, 2007; revised September 06, 2008. First published November 25, 2008; current version published February 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano. This work was supported in part by the Swedish Research Council, Göran Gustafsson Foundation, National Science Foundation, and by the Air Force Office of Scientific Research.

T. T. Georgiou and M. S. Takyar are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: tryphon@ece.umn.edu; shahrouz@ece.umn.edu).

J. Karlsson is with the Department of Mathematics, Division of Optimization and Systems Theory, Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: johan.karlsson@math.kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.2010009

II. Morphisms on Power Spectra

We consider power spectra of discrete-time stochastic processes. These are bounded positive measures on the interval $\mathbb{I} = [-\pi, \pi]$ (with the end points identified) or, in the case of real-valued processes, on $\mathbb{I} = [0, \pi]$ and the set of such measures is denoted by

$$\mathfrak{M} := \{d\mu : d\mu \geq 0 \text{ on } \mathbb{I}\}.$$

The physics of signal interactions suggests certain natural morphisms between spectra that model mixing in the time-domain. The most basic such interactions, additive and multiplicative, adversely affect the information content of signals. It is our aim to devise metrics that respect such a degradation in information content. Another property that ought to be inherent in a metric geometry for power spectra is the continuity of statistics. More specifically, since modeling and identification is often based on statistical quantities, it is natural to demand that "small" changes in the spectral content, as measured by suitable metrics, result in small changes in any relevant statistical quantity.

Consider a discrete-time stationary (in general complex-valued) random process $\{y(k), k \in \mathbb{Z}\}$, or simply y for short, with corresponding power spectrum $d\mu \in \mathfrak{M}$. The sequence of covariances

$$R(\ell) := \mathcal{E}\left\{y(m)\overline{y(m-\ell)}\right\}, \quad \ell = 0, 1, 2, \ldots$$

where $\mathcal{E}\{\cdot\}$ denotes expectation and "$-$" denotes complex conjugation, are the Fourier coefficients of $d\mu$, i.e.

$$R(\ell) = \int_{\mathbb{I}} e^{-j\ell\theta} d\mu(\theta).$$

In general, second-order statistics are integrals of the form

$$\mathbf{R} = \int_{\mathbb{I}} \mathbf{G}(\theta) d\mu(\theta)$$

for an arbitrary vectorial integration kernel $\mathbf{G}(\theta)$ which is continuous in $\theta \in \mathbb{R}$ and periodic with period 2π. For future reference, we denote the set of such functions by $\mathcal{C}_{\text{perio}}(\mathbb{I})$.

Now, suppose that $d\mu_a$ represents the power spectrum of an "additive-noise" process y_a which is independent of y. Then the power spectrum of $y + y_a$ is simply $d\mu + d\mu_a$. Similarly, if $d\mu_m$ represents the power spectrum of a "multiplicative-noise" process y_m, the power spectrum of the point-wise product $y \cdot y_m$ is the circular convolution

$$d\nu = d\mu * d\mu_m,$$

for example, $d\nu$ satisfies

$$\int_{x\in S} d\nu(x) := \int_{x\in S}\int_{t\in\mathbb{I}} d\mu(t)d\mu_a(x-t) \quad \text{for all } S \subseteq \mathbb{I}$$

where the arguments are interpreted modulo 2π.

We postulate situations where we need to discriminate between two signals on the basis of their power spectra and of their statistics. In such cases, additive or multiplicative noise may impede our ability to differentiate between the two. Thus, we consider noise spectra as morphisms on $\mathfrak{M}$ that transform power spectra accordingly. Additive and multiplicative noise morphisms are defined as follows:

$$A_{d\mu_{\rm a}} : d\mu \mapsto d\mu + d\mu_{\rm a}$$

for any $d\mu_a \in \mathfrak{M}$, and

$$M_{d\mu_{\rm m}} : d\mu \mapsto d\mu * d\mu_{\rm m}$$

for any $d\mu_{\rm m} \in \mathfrak{M}$, normalized so that $\int_{\mathbb{I}} d\mu_{\rm m} = 1$. The normalization is such that multiplicative noise is perceived to affect the spectral content but not the total energy of underlying signals.

The effect of additive independent noise on the statistics of a process is also additive, e.g., covariances of the process are transformed according to

$$\hat{A}_{d\mu_{\rm a}} : R(\ell) \mapsto R(\ell) + R_{\rm a}(\ell),$$

where $R_{\rm a}(\ell)$ denotes the corresponding covariances of the noise process. Similarly, multiplicative noise transforms the process statistics by pointwise multiplication (Schur product) as follows:

$$\hat{M}_{d\mu_{\rm m}} : R(\ell) \mapsto R(\ell) \cdot R_{\rm m}(\ell).$$

More generally, $\hat{M}_{d\mu_{\rm m}} : \mathbf{R} \mapsto \mathbf{R} \bullet \mathbf{R}_{\rm m}$ for statistics with respect to an arbitrary kernel $\mathbf{G}(\theta)$, where $\bullet$ denotes point-wise multiplication of the vectors $\mathbf{R}$, $\mathbf{R}_m$.

Consistent with the intuition that noise masks differences between two power spectra, it is reasonable to seek a metric topology, where distances between power spectra are non-increasing when they are transformed by any of the above two morphisms. More precisely, we seek a notion of distance $\delta(\cdot,\cdot)$ on $\mathfrak{M}$ with the following properties:

Axiom i) $\delta(\cdot,\cdot)$ is a metric on $\mathfrak{M}$.

Axiom ii) For any $d\mu_a \in \mathfrak{M}$, $A_{d\mu_{\rm a}}$ is contractive on $\mathfrak{M}$ with respect to the metric $\delta(\cdot,\cdot)$.

Axiom iii) For any $d\mu_m \in \mathfrak{M}$ with $\int_{\mathbb{I}} d\mu_{\rm m} \leq 1$, $M_{d\mu_{\rm m}}$ is contractive on $\mathfrak{M}$ with respect to the metric $\delta(\cdot,\cdot)$.

The property of a map being contractive refers to the requirement that the distance between two power spectra does not increase when the transformation is applied.

An important property for the sought topology of power spectra is that small changes in the power spectra are reflected in correspondingly small changes in statistics. More precisely, any topology induces a notion of convergence, and the question is whether this topology is compatible with the topology of the Euclidean vector-space where (finite) statistics take their values. Continuity of statistics to changes in the power spectra is necessary for quantifying spectral uncertainty based on statistics. The property we require is referred to as (sequential) weak*-continuity and is abstracted in the following statement.

Axiom iv) Let $d\mu \in \mathfrak{M}$ and a sequence $d\mu_k \in \mathfrak{M}$ for $k \in \mathbb{N}$. Then $\delta(d\mu_k, d\mu) \to 0$ as $k \to \infty$, if and only if

$$\int_{\mathbb{I}} \mathbf{G}d\mu_k \to \int_{\mathbb{I}} \mathbf{G}d\mu \text{ as } k \to \infty$$

for any $\mathbf{G} \in \mathcal{C}_{\rm perio}(\mathbb{I})$.

Remark 1: Because the underlying space $\mathbb{I}$ is compact, weak* and sequential weak*-continuity coincide and no distinction will made henceforth.

III. Reflections and Contrast With Information Geometry

The search for natural metrics between density functions can be traced back to the early days of statistics, probability and information theory. According to Chentsov [11, p. 992], [1], Kolmogorov was "always interested in finding *information* distances" between probability distributions. In his notes he emphasized the importance of the total variation

$$d_{\rm TV}(d\mu_0, d\mu_1) := \int |\mu_0(dx) - \mu_1(dx)|$$

as a metric, and he independently arrived at and discussed the relevance of the Bhattacharyya [7] distance

$$d_B(d\mu_0, d\mu_1) := 1 - \int \sqrt{\mu_0(dx)\mu_1(dx)} \tag{1}$$

as a measure of unlikeness of two measures $d\mu_0, d\mu_1$. Both suggestions reveal great intuition and foresight. The total variation admits the following interpretation (cf. [16]) that will turn out to be particularly relevant in our context: the total variation represents the least "energy" of perturbations of two power spectra $d\mu_0$ and $d\mu_1$ that render the two indistinguishable, i.e.

$$d_{\rm TV}(d\mu_0, d\mu_1) = \min\left\{\int d\nu_0 + \int d\nu_1 : d\nu_0, d\nu_1 \in \mathfrak{M}, \right. \\ \left. \text{and } d\mu_0 + d\nu_0 = d\mu_1 + d\nu_1 \right\}. \tag{2}$$

On the other hand the Bhattacharyya distance turned out to have deep connections with Fisher information, the Kullback-Leibler divergence, and the Cramér-Rao inequality. These connections underlie a body of work known as Information Geometry which was advanced by Amari, Nagaoka, Chentsov, and others [3], [12], [22]. At the heart of the subject is the Fisher information metric on probability spaces and the closely related spherical Fisher-Bhattacharyya-Rao metric

$$d_{FBR}(d\mu_0, d\mu_1) := \arccos \int \sqrt{\mu_0(dx)\mu_1(dx)}. \quad (3)$$

This latter metric is precisely the geodesic distance between two distributions in the geometry of the Fisher metric. One of the fundamental results of the subject is Chentsov's theorem. This theorem states that stochastic maps are contractive with respect to the Fisher information metric and moreover, that this metric is in fact the *unique* (up to constant multiple) Riemannian metric with this property [12]. Stochastic maps represent the most general class of linear maps which map probability distributions to the same. Stochastic maps model coarse graining of the outcome of sampling, and thus, form a semi-group. Thus, it is natural to require that any natural notion of distance between probability distributions must be monotonic with respect to the action of stochastic maps.

An alternative justification for the Fisher information metric is based on the Kullback-Leibler divergence

$$\begin{aligned} d_{\mathrm{KL}}(d\mu_0, d\mu_1) &:= \int \frac{d\mu_0}{d\mu_1} \log\left(\frac{d\mu_0}{d\mu_1}\right) d\mu_1 \\ &= \int \log\left(\frac{d\mu_0}{d\mu_1}\right) d\mu_0 \end{aligned}$$

between *probability* distributions. The Kullback-Leibler divergence is not a metric, but quantifies in a very precise sense the difficulty in distinguishing the two distributions [24]. In fact, it may be seen to quantify, in source coding for discrete finite probability distributions, the increase in the average word-length when a code is optimized for one distribution and used instead for encoding symbols generated according to the other. The distance between infinitesimal perturbations, measured using d_{KL}, is precisely the Fisher information metric. It is quite remarkable that both lines of reasoning, degradation of coding efficiency and ability to discriminate on one hand and contractive-ness of stochastic maps on the other, lead to the same geometry on probability spaces.

Turning again to power spectra, we observe that d_{TV} can be used as a metric and has a natural interpretation as explained earlier. The metric d_{FBR} on the other hand can also be used, if suitably modified to account for scaling, but lacks an intrinsic interpretation. A variety of other metrics can also be placed on $\mathfrak{M}$ (cf. [18], [19], [26]), mostly borrowed from functional analysis, which may similarly lack an intrinsic interpretation. Thus, the signal processing community focused instead on other metric-like quantities, as the so-called Itakura-Saito distance [18, Eq. (16)]

$$d_{IS}(f_0 d\theta, f_1 d\theta) := \int \left(\frac{f_0}{f_1} - \ln\left(\frac{f_0}{f_1}\right) - 1\right) d\theta,$$

which have been motivated in the context of linear prediction [18], [27]. The Itakura-Saito distance in particular is intimately related to the probability structure of underlying processes for the Gaussian case and to their distance in the Kullback-Leibler divergence (see, for instance, [23] and [28]). A closely related metric was presented in [14] and [15], which quantifies in a precise way the degradation of predictive error variance --in analogy with the latter argument that led to the Fisher metric. More specifically, a one-step optimal linear predictor for an underlying random process is obtained based on a given power spectrum, and then, this predictor is applied to a random process with a different spectrum. The degradation of predictive error variance, when the perturbations are infinitesimal, gives rise to a Riemannian metric. In this metric, the geodesic distance between two power spectra is

$$d_{\mathrm{pr}}(d\mu_0, d\mu_1) := \sqrt{\int \left(\log \frac{d\mu_0}{d\mu_1}\right)^2 d\theta - \left(\int \log \frac{d\mu_0}{d\mu_1} d\theta\right)^2} \quad (4)$$

which effectively depends on the ratio of the corresponding spectral densities. Interestingly, this metric is a "normalized version" of the so-called log-spectral distance [26, Eq. (78)]

$$d_{LS}(d\mu_0, d\mu_1) := \sqrt{\int \left(\log \frac{d\mu_0}{d\mu_1}\right)^2 d\theta}$$

which is commonly used without any intrinsic justification. A similar rationale that leads to (4) can be based on degradation of smoothing-variance instead of prediction (see [14] and [15]), and this also leads to expressions that weigh in ratios of the corresponding spectral density functions, i.e., it is the ratios of the absolutely continuous part of the measures that play any role.

A possible justification for such metrics which weigh in only the ratio of the corresponding density functions can be sought in interpreting the effect of linear filtering as a kind of processing that needs to be addressed in the axioms. More specifically, the power spectrum at the output of a linear filter relates to the power spectrum of the input via multiplication by the modulus square of the transfer function. Thus, a metric that respects such "processing" ought to be contractive (and possibly invariant). However, it turns out that such a property is incompatible with the spectral properties that we would like to have, and in particular it is incompatible with the ability of the metric to localize a measure based on its statistics (cf. Axiom iv)). This incompatibility is shown next.

Consider morphisms on $\mathfrak{M}$ that correspond to processing by a linear filter

$$F_h : d\mu \mapsto |h|^2 d\mu$$

for any $h \in H_\infty$. Here, h is thought of as the transfer function of the filter, μ the power spectrum of the input, and $|h|^2 d\mu$ the power spectrum of the output.

Proposition 2: Assume that $\delta(\cdot,\cdot)$ is a weak*-continuous metric on $\mathfrak{M}$. Then there exists $h \in H_\infty$ such that F_h is not contractive with respect to $\delta(\cdot,\cdot)$.

Proof: We will prove the claim by showing that whenever δ is a weak*-continuous metric that satisfies Property i), we may derive a contradiction. Denote by μ_t, $t \geq 0$ the measure with a unit mass in the point t and let $\epsilon = \delta(d\mu_0, d\mu_0/2)$. By weak*-continuity, there exists $t_0 > 0$ such that $\delta(d\mu_0, d\mu_{t_0}) <$

$\epsilon/3$. Let $h \in H_\infty$ be such that $|h(0)|^2 = 1/2$ and $|h(t_0)|^2 = 1$. Then we have that

$$\begin{aligned}\epsilon &= \delta(d\mu_0, d\mu_0/2) \leq \delta\left(d\mu_0, d\mu_{t_0}\right) + \delta\left(d\mu_{t_0}, d\mu_0/2\right)\\ &= \delta\left(d\mu_0, d\mu_{t_0}\right) + \delta\left(|h|^2\mu_{t_0}, |h|^2\mu_0\right)\\ &\leq 2\delta\left(d\mu_0, d\mu_{t_0}\right) < \frac{2}{3}\epsilon.\end{aligned}$$

Which is a contradiction, and, hence, the proposition holds. ■

It is important to point out that none of the above [i.e., neither (3) nor (4)] is a weak*-continuous metric. In particular, the metric in (4) is impervious to spectral lines as only the absolutely continuous part of the spectra plays any role. Similarly, neither the metric in (2) nor the one in (3) can localize distributions because they are not weak*-continuous. Thus, in this paper, we follow a line of reasoning analogous to the axiomatic framework of the Chentsov theorem, but for power spectra, requiring a metric to satisfy Axioms i)–iv).

IV. The Monge-Kantorovich Problem

A natural class of metrics on measures are transport metrics based on the ideas of Monge and Kantorovich. The Monge-Kantorovich distance represents a cost of moving a nonnegative measure $d\mu_0 \in \mathfrak{M}(X)$ to another nonnegative measure $d\mu_1 \in \mathfrak{M}(X)$, given that there is an associated cost $c(x,y)$ of moving mass from the point x to the point y. The theory may be formulated for rather general spaces X, but in this paper we restrict our attention to compact metric spaces X and, in particular, $\mathbb{I}$. Every possible way of moving the measure $d\mu_0$ to $d\mu_1$ corresponds to a transference plan $\pi \in \mathfrak{M}(X\times X)$, which satisfies

$$\int_{y\in X} d\pi(x,y) = d\mu_0 \text{ and } \int_{x\in X} d\pi(x,y) = d\mu_1$$

or more rigorously, that

$$\pi[A\times X] = \mu_0(A) \text{ and } \pi[X\times B] = \mu_1(B) \qquad (5)$$

whenever $A, B \subset X$ are measurable. Such a plan exists only if the measures $d\mu_0$ and $d\mu_1$ have the same mass, i.e., $\mu_0(X) = \mu_1(X)$. Denote by $\Pi(d\mu_0, d\mu_1)$ the set of all such transference plans, i.e.,

$$\Pi(d\mu_0, d\mu_1) = \{\pi \in \mathfrak{M}(X\times X) : (5) \text{ holds for all } A, B\}.$$

To each such transference plan, the associated cost is

$$\mathcal{I}[\pi] = \int_{X\times X} c(x,y)\pi(x,y)$$

and consequently, the minimal transportation cost is

$$T_c(d\mu_0, d\mu_1) := \min\left\{\mathcal{I}(\pi) : \pi \in \Pi(d\mu_0, d\mu_1)\right\}. \qquad (6)$$

The optimal transportation problem admits a dual formulation, referred to as the Kantorovich duality, which we state below without a proof. For a derivation and an insightful exposition we refer the reader to [32, p. 19].

Theorem 3: Let c be a lower semicontinous (cost) function, let

$$\Phi_c := \left\{(\phi,\psi) \in L^1(d\mu_0)\times L^1(d\mu_1) : \phi(x)+\psi(y) \leq c(x,y)\right\}$$

and let

$$\mathcal{J}(\phi,\psi) = \int_X \phi d\mu_0 + \psi d\mu_1.$$

Then

$$T_c(d\mu_0, d\mu_1) = \sup_{(\phi,\psi)\in\Phi_c} \mathcal{J}(\phi,\psi).$$

A rather simple consequence is the following lemma.

Lemma 4: Let c be a lower semicontinous (cost) function with $c(x,x) = 0$ for $x \in X$. Then $A_{d\mu_a}$ is contractive with respect to T_c.

Proof: Contractiveness of $A_{d\mu_a}$ follows from the dual representation. Any pair $(\phi,\psi) \in \Phi_c$ satisfies $\phi(x) + \psi(x) \leq 0$, and, hence

$$\int_X \phi d\mu_0 + \psi d\mu_1 \geq \int_X \phi d\mu_0 + \psi d\mu_1 + (\phi+\psi)d\mu_a.$$

■

Monge-Kantorovich distances are not metrics, in general, but they readily give rise to a class of the so-called Wasserstein metrics as explained next.

Theorem 5: Assume that the (cost) function $c(\cdot,\cdot)$ is of the form $c(x,y) = d(x,y)^p$ where d is a metric and $p \in (0,\infty)$. Then the Wasserstein distance

$$W_p(d\mu_0, d\mu_1) = T_c(d\mu_0, d\mu_1)^{\min\left(1,\frac{1}{p}\right)}$$

is a metric on the subspace of $\mathfrak{M}(X)$ with fixed mass and metrizes the weak* topology.

Proof: See [32, Ch. 7]. Note that since X is compact, the weak* topology on $\mathfrak{M}(X)$ coincides with the weak topology. ■

V. Metrics Based on Transportation

The Monge-Kantorovich theory deals with measures of equal mass. As we have just seen, it provides metrics that have some of the properties that we seek to satisfy. The purpose of this section is to develop a metric based on similar principles, that applies to measures of possibly unequal mass.

Given two nonnegative measures $d\mu_0$ and $d\mu_1$ on $\mathbb{I}$, we postulate that these are perturbations of two other measures $d\nu_0$ and $d\nu_1$, respectively, which have equal mass. Then, the cost of transporting $d\mu_0$ and $d\mu_1$ to one another can be thought of as the cost of transporting $d\nu_0$ and $d\nu_1$ to one another plus the size of the respective perturbations. Thus we define

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) := \inf_{\nu_0(\mathbb{I})=\nu_1(\mathbb{I})} T_c(d\nu_0, d\nu_1) + \kappa\sum_{i=0}^{1} d_{\mathrm{TV}}(d\mu_i, d\nu_i) \qquad (7)$$

where κ is a suitable parameter that weighs the relative contribution of perturbation and transportation. Define

$$c(x,y) = |(x-y)_{\text{mod } 2\pi}|^p \tag{8}$$

where $(x)_{\text{mod}2\pi}$ is the element in the equivalence class $x+2\pi\mathbb{Z}$ which belongs to $(-\pi,\pi]$. The main result of the section is the following theorem.

Theorem 6: Let $\kappa > 0$ and $c(x,y)$ defined as in (8), where $p \in (0,\infty)$. Then

$$\delta_{p,\kappa}(d\mu_0, d\mu_1) := \left(\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)\right)^{\min\left(1,\frac{1}{p}\right)}$$

is a metric on $\mathfrak{M}$ which satisfies Axiom i)–iv).

The proof uses the fact that (7) has an equivalent formulation as a transportation problem, and a corresponding dual stated below.

Theorem 7: Let c be a lower semicontinuous (cost) function, let

$$\Phi_{c,\kappa} := \{(\phi,\psi) \in L^1(d\mu_0) \times L^1(d\mu_1) : \\ \phi(x) \le \kappa, \psi(y) \le \kappa, \phi(x)+\psi(y) \le c(x,y)\}$$

and let

$$\mathcal{J}(\phi,\psi) = \int_{\mathbb{I}} \phi d\mu_0 + \psi d\mu_1.$$

Then

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \mathcal{J}(\phi,\psi). \tag{9}$$

Remark 8: Definition (7) does not provide a direct way to compute $\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)$, whereas the dual formulation in Theorem 7 is amenable to numerical implementation. Indeed, (9) is a linear optimization problem which can be computed using standard methods.

Proof: The problem (7) can be thought of as a transportation problem on the set $X = \mathbb{I} \cup \{\infty\}$, where a mass is added at ∞ as needed to normalize the measures so that they have equal mass, e.g.

$$\hat{\mu}_i(S) = \mu_i(S) \text{ for } S \subset \mathbb{I}$$
$$\hat{\mu}_i(\infty) = M - \mu_i(\mathbb{I})$$

for some $M \ge \max\{\mu_i(\mathbb{I}) : i = 0,1\}$. Accordingly, the (cost) function is modified as follows:

$$\hat{c}(x,y) = \begin{cases} \min(c(x,y), 2\kappa) & \text{for } x,y \in \mathbb{I}, \\ \kappa & \text{for } x \in \mathbb{I}, y = \infty, \\ \kappa & \text{for } x = \infty, y \in \mathbb{I}, \\ 0 & \text{for } x = \infty, y = \infty. \end{cases} \tag{10}$$

First we prove that

$$T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) = \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \mathcal{J}(\phi,\psi),$$

hence $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1)$ is independent of M, and then we conclude the proof by showing that

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1). \tag{11}$$

According to Theorem 3, $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1)$ is equal to the supremum of

$$\hat{\mathcal{J}}(\phi,\psi) := \int_X \phi d\hat{\mu}_0 + \psi d\hat{\mu}_1$$

subject to

$$\phi(x) + \psi(y) \le \hat{c}(x,y) \quad \text{for} \quad x,y \in \mathbb{I}, \tag{12}$$
$$\phi(x) + \psi(\infty) \le \kappa \quad \text{for} \quad x \in \mathbb{I}, \tag{13}$$
$$\phi(\infty) + \psi(y) \le \kappa \quad \text{for} \quad y \in \mathbb{I}, \tag{14}$$
$$\phi(\infty) + \psi(\infty) \le 0. \tag{15}$$

Our first claim now follows by showing that there is no added restriction imposed by requiring that $\phi(\infty) = \psi(\infty) = 0$. Indeed, $\Phi_{c,\kappa}$ is essentially identical to the set

$$\{(\phi,\psi) : (12)-(15) \text{ hold and } \phi(\infty) = \psi(\infty) = 0\}$$

with (ϕ,ψ) extended to have support at ∞ as well. To this end, let (ϕ,ψ) be an arbitrary pair of functions satisfying (12)–(15). Since additive scaling of $(\phi,-\psi)$ does not change the constraints nor the value of $\hat{\mathcal{J}}(\phi,\psi)$, we may assume that $\phi(\infty) = 0$. There are two cases that we need to consider. If $\sup_{x\in\mathbb{I}} \phi(x) \le \kappa$, then define

$$\hat{\phi}(x) = \phi(x) \text{ for } x \in X, \hat{\psi}(x) = \begin{cases} \psi(x) & x \in \mathbb{I} \\ 0 & x = \infty, \end{cases}$$

and if $\sup_{x\in\mathbb{I}} \phi(x) = \epsilon + \kappa > \kappa$, then define

$$\hat{\phi}(x) = \begin{cases} \phi(x) - \epsilon & x \in \mathbb{I} \\ 0 & x = \infty \end{cases}, \quad \hat{\psi}(x) = \begin{cases} \psi(x) + \epsilon & x \in \mathbb{I} \\ 0 & x = \infty. \end{cases}$$

In both cases, we have that $\hat{\mathcal{J}}(\phi,\psi) \le \hat{\mathcal{J}}(\hat{\phi},\hat{\psi})$ as well as that the pair $(\hat{\phi},\hat{\psi})$ satisfies (12)–(15). In the second case, the constraint (14) is not violated; $\hat{c}(x,y) \le 2\kappa$ implies that $\sup_{x\in\mathbb{I}} \phi(x) + \sup_{y\in\mathbb{I}} \psi(y) \le 2\kappa$, and hence $\sup_{y\in\mathbb{I}} \psi(y) \le \kappa - \epsilon$. Note that (13) implies that $\psi(\infty) \le -\epsilon$. Thus, in both cases, from an arbitrary pair (ϕ,ψ), we have constructed a pair $(\hat{\phi},\hat{\psi})$ for which the constraints (12)–(15) hold, $\hat{\phi}(\infty) = \hat{\psi}(\infty) = 0$ holds, and the value of $\hat{\mathcal{J}}(\phi,\psi)$ has not decreased. Therefore we may without loss of generality let $\phi(\infty) = \psi(\infty) = 0$.

It remains to show that $\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1)$. We start by showing that $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) \ge \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)$. Let $\hat{\pi} \in \Pi(d\hat{\mu}_0, d\hat{\mu}_1)$ and let measures ν_0 and ν_1 be defined via

$$\int_{y\in\mathbb{I}, x\in S} d\hat{\pi}(x,y) = \nu_0(S) \text{ and } \int_{x\in\mathbb{I}, y\in S} d\hat{\pi}(x,y) = \nu_1(S),$$

for $S \subseteq \mathbb{I}$. Then for any transference plan $\hat{\pi}$ we have that

$$\begin{aligned}\mathcal{I}[\hat{\pi}] &= \int_{\mathbb{I}\times\mathbb{I}} \hat{c}(x,y)\hat{\pi}(x,y) + \kappa \int_{\mathbb{I}} d\hat{\pi}(x,\infty) + \kappa \int_{\mathbb{I}} d\hat{\pi}(\infty,y) \\ &= \int_{\mathbb{I}\times\mathbb{I}} \hat{c}(x,y)\hat{\pi}(x,y) + \kappa \sum_{i=0}^{1} \int_{\mathbb{I}} (d\mu_i - d\nu_i) \\ &\geq T_c(d\nu_0, d\nu_1) + \kappa \sum_{i=0}^{1} d_{\mathrm{TV}}(d\mu_i, d\nu_i) \\ &\geq \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1).\end{aligned}$$

To see this, note that $d\mu_i - d\nu_i$ is positive and, hence, that $\int_{\mathbb{I}}(d\mu_i - d\nu_i) = d_{\mathrm{TV}}(d\mu_i, d\nu_i)$. Therefore, $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) \geq \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)$ always holds. To show that the reverse inequality also holds let $d\nu_0, d\nu_1$ be two nonnegative measures with $\nu_0(\mathbb{I}) = \nu_1(\mathbb{I})$. By introducing $f_0 = \max(-\kappa, \phi)$, $f_1 = \max(-\kappa, \psi)$ and using the dual formulation we get

$$\begin{aligned}T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) &= \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \mathcal{J}(\phi,\psi) \\ &\leq \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \int_{\mathbb{I}} f_0 d\mu_0 + f_1 d\mu_1 \\ &\leq \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \sum_{i=0}^{1} \int_{\mathbb{I}} f_i d\nu_i + f_i(d\mu_i - d\nu_i) \\ &\leq \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \sum_{i=0}^{1} \int_{\mathbb{I}} f_i d\nu_i + \kappa \sum_{i=0}^{1} d_{\mathrm{TV}}(d\mu_i, d\nu_i) \\ &\leq T_c(d\nu_0, d\nu_1) + \kappa \sum_{i=0}^{1} d_{\mathrm{TV}}(d\mu_i, d\nu_i).\end{aligned}$$

Here, we use the fact that $\kappa d_{\mathrm{TV}}(d\mu_i, d\nu_i) = \sup_{\|f\|_\infty \leq \kappa} \int f(d\mu_i - d\nu_i)$. Since the above inequality holds for any measures $d\nu_0, d\nu_1$ we get the reversed inequality $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) \leq \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)$, and hence we conclude that $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) = \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)$. ■

The final step to proving Theorem 6 will be provided by the following lemma.

Lemma 9: Let $c(x,y)$ be a function of $|x-y|$. Then for any $d\mu_{\mathrm{m}} \in \mathfrak{M}$ with $\int_{\mathbb{I}} d\mu_{\mathrm{m}} \leq 1$, $M_{d\mu_{\mathrm{m}}}$ is contractive on $\mathfrak{M}$ with respect to $\tilde{T}_{c,\kappa}$.

Proof: Note that

$$\begin{aligned}&\int_{x\in\mathbb{I}} \phi(x)(d\mu_m * d\mu_0)(x) \\ &\quad= \int_{x\in\mathbb{I}} \phi(x) \int_{\tau\in\mathbb{I}} d\mu_m(x-\tau)d\mu_0(\tau) \\ &\quad= \int_{\tau\in\mathbb{I}} \left(\int_{x\in\mathbb{I}} \phi(x)d\mu_m(x-\tau) \right) d\mu_0(\tau) \\ &\quad= \int_{\tau\in\mathbb{I}} \left(\phi(x) * d\mu_m(-x)\right)|_\tau \, d\mu_0(\tau)\end{aligned}$$

and denote

$$\begin{aligned}\phi_m(\tau) &= \phi(x) * d\mu_m(-x)|_\tau \\ \psi_m(\tau) &= \psi(x) * d\mu_m(-x)|_\tau.\end{aligned}$$

From this it follows that

$$\mathcal{J}_{(d\mu_m * d\mu_0, d\mu_m * d\mu_1)}(\phi,\psi) = \mathcal{J}_{(d\mu_0, d\mu_1)}(\phi_m, \psi_m)$$

where the subscript specifies the measures used in the definition of the dual functional.

Now let $(\phi,\psi) \in \Phi_{c,\kappa}$. Then

$$\begin{aligned}\phi(x-\tau) + \psi(y-\tau) &\leq \min\left(c(x-\tau, y-\tau), 2\kappa\right) \\ &= \min\left(c(x,y), 2\kappa\right)\end{aligned}$$

and by integrating with respect to $d\mu_m(-\tau)$ over $\tau \in \mathbb{I}$, we arrive at

$$\phi_m(x) + \psi_m(y) \leq \min\left(c(x,y), 2\kappa\right).$$

Furthermore, it is immediate that $\phi(x) \leq \kappa$ and $\psi(y) \leq \kappa$ implies that $\phi_m(x) \leq \kappa$ and that $\psi_m(y) \leq \kappa$, and, hence, $(\phi_m, \psi_m) \in \Phi_{c,\kappa}$ follows. Finally

$$\begin{aligned}&\tilde{T}_{c,\kappa}\left(M_{d\mu_m}(d\mu_0), M_{d\mu_m}(d\mu_1)\right) \\ &\quad= \sup_{(\phi,\psi)\in\Psi_{c,\kappa}} \mathcal{J}_{(d\mu_m * d\mu_0, d\mu_m * d\mu_1)}(\phi,\psi) \\ &\quad= \sup_{(\phi,\psi)\in\Psi_{c,\kappa}} \mathcal{J}_{(d\mu_0, d\mu_1)}(\phi_m,\psi_m) \\ &\quad\leq \sup_{(\phi,\psi)\in\Psi_{c,\kappa}} \mathcal{J}_{(d\mu_0, d\mu_1)}(\phi,\psi) \\ &\quad= \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)\end{aligned}$$

■

We now recap the proof of our main theorem.

Proof: [Proof of theorem 6]: From (11), $\tilde{T}_{c,\kappa}$ can be viewed as the cost of a transportation problem. Since the associated cost function $\hat{c}$ from (10) is of the form d^p, where d is a metric, Axiom i) follows from Theorem 5. Axiom iv) follows by noting that if a sequence of measures $d\mu_n$ converges to $d\nu$ in weak*, then in particular $\mu_n(\mathbb{I}) \to \nu(\mathbb{I})$. Therefore, for any $M > \sup\{\mu_n(\mathbb{I}), n \in \mathbb{N}, \text{and } \nu(\mathbb{I})\}$, it readily follows that $\hat{\mu}_n$ converges to $\hat{\nu}$ in weak*, and, hence, Theorem 5 ensures weak*-continuity of $\delta_{p,\kappa}$. From the above formulation, Axiom ii) follows from Lemma 4. Finally Axiom iii) follows from Lemma 9. ■

Remark 10: It is interesting to note that for the case $p = 1$

$$\delta_{1,\kappa}(d\mu_0, d\mu_1) = \max_{\substack{\|g\|\leq\kappa \\ \|g\|_L \leq 1}} \int g(d\mu_0 - d\mu_1)$$

where $\|f\|_L = \sup(|f(x) - f(y)|/|x-y|)$ the Lipschitz norm. Furthermore, in general, for any p

$$\frac{1}{\kappa}\delta_{1,\kappa}(d\mu_0, d\mu_1) \to d_{\mathrm{TV}}(d\mu_0, d\mu_1) \text{ as } \kappa \to 0.$$

Remark 11: The transportation problem has indeed some very nice properties that relate to the weak*-continuity of the corresponding distance metrics. In particular, smoothness with respect to translations and small deformations is an intrinsic

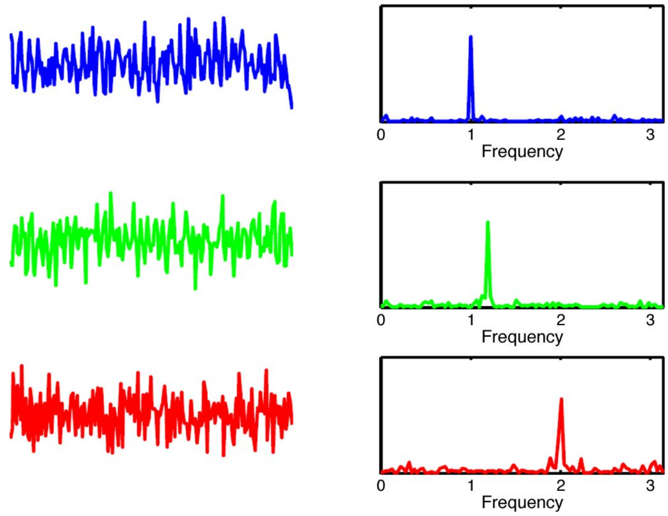

Fig. 1. Stochastic process y_k in time and frequency domain for $\theta = 1$, 1.2, and 2.

TABLE I
COMPARISON OF DISTANCE MEASURES ON SPECTRAL LINES + NOISE

	Distance between line spectra at		
	$\theta_1 = 1$ $\theta_2 = 1.2$	$\theta_2 = 1.2$ $\theta_3 = 2$	$\theta_1 = 1$ $\theta_3 = 2$
Transportation: $\delta_{1,1}$	0.3077	0.8832	1.0690
Predictive: d_{pr}	1.8428	1.8390	1.8517
Itakura Saito: d_{IS}	22.7279	472.6690	134.1707

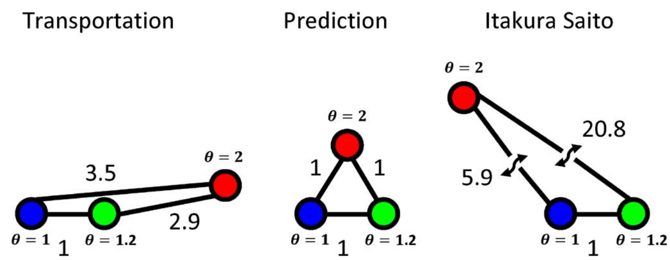

Fig. 2. Relative distances between line spectra.

property. For this reason, it has been used in conjunction with other notions of distance, in a similar fashion as in the current paper, to link density functions of unequal mass. In particular, Benamou and Brenier [6] have introduced a mixed L^2/Wasserstein optimal mapping to link such density functions, while in other relevant literature, Caffarelli and McCann [10] and more recently, Figalli [13] study the transportation of a portion of two unequal masses onto each other.

VI. EXAMPLES

We present two examples that highlight the relevance of the proposed metrics in spectral analysis. The first example compares how different distance measures perform on spectra which contain spectral lines. The second compares how these measures distinguish voiced sounds of different speakers. The distance measures we consider, besides the transportation distance (here $\delta_{1,1}$), are the prediction metric and the Itakura-Saito distance. In both examples the time-series are normalized to have the same variance.

Example 12: We consider a random process $y_k = \cos(k\theta + \phi) + w_k$ which consists of a sinusoidal component and a zero-mean, unit-variance, white-noise component w_k. Here, θ is taken as a constant, whereas ϕ is assumed random, independent of w_k, and uniformly distributed on $(-\pi, \pi]$. Fig. 1 shows three samples of such a random process for respective values of $\theta \in \{1, 1.2, 2\}$, along with their respective power spectra. Based on a set of 500 independent simulations, Table I shows the average distance of the respective power spectra when measured using i) the transport distance, ii) the prediction distance [15], and iii) the Itakura-Saito distance (see, e.g., [18]). Comparison of these values reveals that only the transportation-based metric can reliably distinguish between spectral lines.

The schematic in Fig. 2 compares the relative distances in these three cases with the smallest value normalized to one. The respective distances for the case of the prediction metric are relatively insensitive to the actual location of the spectral line, as in the limit of a long observation record all three distances ought to be equal to one. Recall that the prediction metric does not detect deterministic components. On the other hand the Itakura-Saito distance gives a rather distorted view of reality. In the transportation metric, the respective distances are consistent with "physical" location of the spectral lines. Further, the consistency in the ability to discriminate between such spectra is dramatically different in the three cases. Consider the proportion of the simulations for which the distance between the first two spectra ($\theta = 1, 1.2$) is smaller than any of the other distances ($\theta = 1.2, 2$ or $\theta = 1, 2$). For the transport distance in all 500 iterations the distance between the first two power spectra with lines at $\{1, 1.2\}$ was smaller than the distance between the other two possibilities. On the other hand, the corresponding percentages for the prediction distance and for the Itakura-Saito distance were 34.4% and 33.8%, respectively. Thus, the transportation correctly identifies the two spectra that are intuitively closest (i.e., having spectral lines closest to each other), whereas the other distance measures succeed about one third of the times (practically a random pick).

To be fair, neither the prediction metric nor the Itakura-Saito distance were designed, or claimed, to have such discrimination capabilities. As the sample size tends to infinity power spectra computed via the periodogram method converge to the true spectrum in weak*, and since the transportation distance is weak*-continuous, transportation distances converge to the true values. On the other hand, this is not the case for either of the other two distances. □

Example 13: Fig. 3 (left side column) shows time samples corresponding to the phoneme "a" spoken by three individuals, speakers A (Alice), B (Bob), and C (Colin), respectively. Speakers B and C were chosen to be males whereas speaker A was chosen to be a female and, accordingly, the dominant formant of the first speaker has higher pitch than the other two,

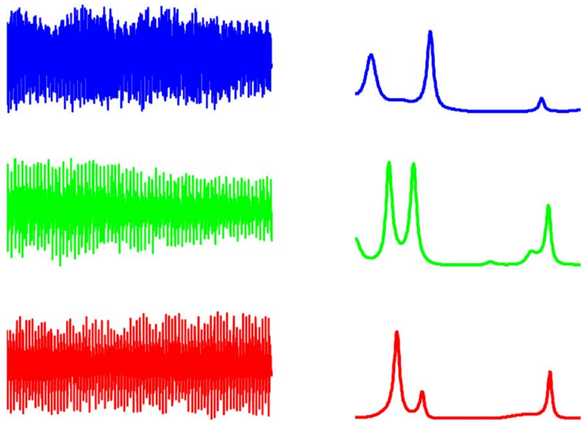

Fig. 3. Phoneme "a" for Alice, Bob, and Colin (top to bottom) in time and frequency domain.

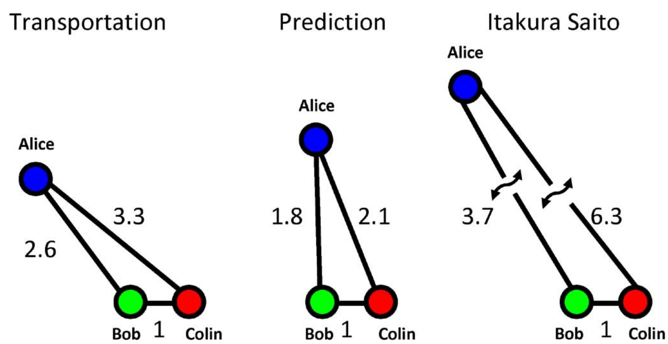

Fig. 4. Relative distances between spectra of the voiced sounds in Fig. 3.

as seen in the estimated power spectra shown in Fig. 3 (right hand side column). The distance between these three power spectra in the transportation metric, the prediction metric, and the Itakura-Saito distance are compared in Fig. 4 as before. The shortest distance in all three cases is normalized to one.

It is seen that all three distance measures are consistent in that the power spectrum corresponding to Bob is always between that of Alice and Colin. However, the respective distances are highly skewed, especially when it comes to the Itakura-Saito distance. □

In the second example, all three distance measures appear to give qualitatively similar results. However, in general, visual comparison of two power spectra appears to be more easily correlated with respective distances in the transportation metric. This is rather evident in the first example and the reasons are traceable to the physical interpretation of the metric with regard to mass transfer. Moreover, we should point out that in neither example did the Itakura-Saito distance respect the triangular inequality (and of course, it has never been claimed to satisfy this metric property).

VII. Concluding Remarks

Our goal has been to identify natural notions of distance for quantitative spectral analysis. Historically, there has been a variety of options [18], [19], [26], [27] which were used to measure distortion, and which were motivated by their perceptive qualities, e.g., see [4] and [30]. However, in order to quantify spectral uncertainty, the relevant metrics ought to allow localization of power spectra based on estimated statistics (i.e., being weak^* continuous). At the same time, these metrics ought to share certain natural properties with regard to how noise affects distance between power spectra. In the present paper, we have presented an axiomatic framework that attempts to capture these intuitive notions and we have developed a family of metrics that satisfy the stated requirements.

While there are many possibilities for developing weak^*-continuous metrics as suggested, we have chosen to base our approach on the concept of transportation. The reason is that the resulting metrics have certain additional properties which relate to deformations of spectra and smoothness with respect to translation. More specifically, from experience, it appears that geodesics (in, e.g., the Wasserstein 2-metric) preserve "lumpiness." A consequence is that when linking power spectra of two similar speech sounds via geodesics of the metric, the corresponding formants often seem to be "matched" and the power between those to transfer in a consistent manner. Such a property appears highly desirable in speech morphing (cf., see [21]). Thus, it will be interesting to expand the set of axioms to include such desirable properties in a more formal way. Such an exercise may in turn narrow down the possible choices of metrics and provide further justification for transportation-based metrics.

Finally, we wish to comment on the need for analogous metrics for comparing multivariable spectra. The ability to localize matricial power spectra is of great significance in system identification, as for instance, in identification based on joint statistics of the input and output processes of a system (cf. [17]).

References

[1] B. P. Adhikari and D. D. Joshi, "Distance, discrimination et resumé exaustif," *Publ. Inst. Univ. Paris*, vol. 5, pp. 57–74, 1956.
[2] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Royal Stat. Soc.*, vol. 28, pp. 131–142, 1966.
[3] S. Amari and H. Nagaoka*, Methods of Information Geometry* Transl.:Translations of Mathematical Monographs, AMS. Oxford, U.K.: Oxford Univ. Press, 2000.
[4] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Process.*, vol. 18, no. 4, pp. 349–369, Dec. 1989.
[5] E. F. Beckenbach and R. Bellman*, Inequalities*. Berlin-Heidelberg, Germany: Springer-Verlag, 1965, 198 pp..
[6] J. D. Benamou and Y. Brenier, "Mixed L^2/Wasserstein optimal mapping between prescribed density functions," *J. Optimiz. Theory Appl.*, vol. 111, no. 2, pp. 255–271, Nov. 2001.
[7] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.
[8] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. Math. Phys.*, vol. 7, pp. 200–217, 1967.
[9] A. Buzo, F. Kuhlmann, and C. Rivera, "Rate-distortion bounds for quotient-based distortions with appolication to Itakura-Saito distortion measures," *IEEE Trans. Inf. Theory*, vol. 32, no. 2, pp. 141–147, Mar. 1986.
[10] L. A. Caffarelli and R. J. McCann, "Free boundaries in optimal transport and Monge-Ampère obstacle problems," *Ann. Math.* [Online]. Available: http://www.math.toronto.edu/mccann/papers/free.pdf, to be published
[11] N. N. Chentsov, "The unfathomable influence of Kolmogorov," *The Ann. Statist.*, vol. 18, no. 3, pp. 987–998, 1990.
[12] N. N. Chentsov*, Statistical Decision Rules and Optimal Inference* Transl.:English translation, Providence. Moscow, Russia: Nauka, 1982.
[13] A. Figalli, The Optimal Partial Transport Problem Univ. Nice-Sophia Antipolis, preprint, April 2008.

[14] T. T. Georgiou, "Distances between power spectral densities," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 3993–4003, Aug. 2007.
[15] T. T. Georgiou, "An intrinsic metric for power spectral density functions," *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 561–563, Aug. 2007.
[16] T. T. Georgiou, "Distances between time-series and their autocorrelation statistics," in *Modeling, Estimation and Control*, A. Chiuso, A. Ferrante, and S. Pinzoni, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 113–122, Festschrift in Honor of Giorgio Picci.
[17] T. T. Georgiou, C. R. Shankwitz, and M. C. Smith, "Identification of linear systems: A graph point of view," in *Amer. Control Conf.*, 1982.
[18] R. Gray, A. Buzo, A. Gray, and Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, Aug. 1980.
[19] A. H. Gray, Jr. and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 380–391, Oct. 1976.
[20] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, 2nd ed. New York: Chelsea, 1984.
[21] X. Jiang, S. Takyar, and T. T. Georgiou, "Metrics and morphing of power spectra," in *Lecture Notes in Control and Information Sciences, Recent Advantages in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds. New York: Springer Verlag, 2008, vol. 371.
[22] R. E. Kass, "The geometry of asymptotic inference," *Statist. Sci.*, vol. 4, no. 3, pp. 188–234, 1989.
[23] L. Knockaert, "A class of statistical and spectral distance measures based on Bose-Einstein statistics," *IEEE Trans. Signal Process.*, vol. 41, no. 11, pp. 3174–3177, Nov. 1993.
[24] S. Kullback, *Information Theory and Statistcs*. New York: Dover, 1997.
[25] J. Karlsson and T. T. Georgiou, "Signal analysis, moment problems and uncertainty measures," in *Proc. IEEE Int. Conf. Decision Control*, Dec. 2005, pp. 5710–5715.
[26] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.
[27] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
[28] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. New York: Holden-Day, 1964.
[29] S. Rachev and L. Rüschendorf, *Mass Transportation Problems, Volume 1: Theory*. New York: Springer, 1998.
[30] F. Soong and M. M. Sondhi, "A frequency-weighted Itakura spectral distortion measure and its application to speech recognition in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-36, no. 1, pp. 41–48, Jan. 1988.
[31] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2005.
[32] C. Villani, *Topics in Optimal Transportation*. New York: AMS, 2003, vol. 58, Graduate studies in Mathematics.

Tryphon T. Georgiou (M'79–SM'99–F'00) was born in Athens, Greece, on October 18, 1956. He received the Diploma in mechanical and electrical engineering from the National Technical University of Athens, in 1979, and the Ph.D. degree from the University of Florida, Gainesville, in 1983.

From 1983–1986, he served on the faculty of Florida Atlantic University, Boca Raton, and from 1986–1989 was with Iowa State University, Ames. Since 1989, he has been with the University of Minnesota, Minneapolis, where he holds the Vincentine Hermes-Luh Chair of Electrical Engineering.

Dr. Georgiou has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL (1991–1992), the *SIAM Journal on Control and Optimization* (1988–1995), the *Systems and Control Letters* (1995–present), and the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2007–present). He has served on the Board of Governors of the Control Systems Society of the IEEE (2002–2005). He has been a recipient of the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society three times, for the years 1992, 1999, and 2003. In 1992 and 1999, he received the award for joint work with Prof. M. C. Smith (Cambridge Univ., U.K.), and in 2003, for joint work with Professors C. Byrnes of Washington University, St. Louis, MO, and A. Lindquist of the Royal Institute of Technology (KTH), Stockholm, Sweden.

Johan Karlsson was born in Stockholm, Sweden, 1979. He received the M.S. degree in engineering physics from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2003. He spent the academic year 2000–2001 as an exchange student at Washington University, St. Louis, MO, and did his master thesis at the University of Minnesota, Minneapolis.

Since fall 2003, he has been a graduate student at the Division of Optimization and Systems Theory, KTH. His research interests includes analytic interpolation, distance measures, and model reduction for applications in signal processing and control theory.

Mir Shahrouz Takyar (S'05) received the B.Sc. and M.Sc. degrees in electrical engineering from Sharif University of Technology, Tehran, in 2000 and 2002, respectively. He also received a second M.Sc. degree from the Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, in 2007.

Since spring 2003, he has been working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota. His current research interests include the areas of control theory and signal processing, especially analytic interpolation and its relevance in engineering disciplines such as circuit theory, stochastic processes, and optimization.