Metrics for power spectra: an axiomatic approach

Tryphon T. Georgiou, Johan Karlsson, and Mir Shahrouz Takyar

***Abstract*— We present an axiomatic framework for seeking a distance 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 the require continuity of statistical quantities with respect to perturbations measured in the metric. The purpose of this paper is to explore certain notions of distance which are based on the Monge-Kantorovic transportation problem and satisfy the natural set of axioms that we have put forth. These type of distance measures are contrasted with an earlier Riemannian metric which was motivated by the geometry of the underlying time-series.**

***Index Terms*— Power spectra, spectral distances, metrics, geodesics, geometry of spectral measures.**

EDICS category: SSP-SPEC

Corresponding author: T.T. Georgiou
Department of Electrical and Computer Engineering
200 Union street, S.E., EECSci 5-179
University of Minnesota, Minneapolis, MN 55455

Email. tryphon@umn.edu
Tel. (612) 625-3303, Fax. (612) 625-4583

I. Introduction

A key element of any quantitative scientific theory is a well-defined and natural metric. A model for the development of such a metric 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 any statistic is continuous with respect to spectral uncertainty quantified by the sought metric. We build on [14] where a variety of metrics were studied, based on complex analysis, which quantify spectral uncertainty given estimated statistics. The focus of the current paper is twofold, firstly to develop a natural axiomatic framework for seeking such geometries for power spectra, and secondly to propose a particular candidate metric which abides by the axioms. This metric is based on the Monge-Kantorovic transportation problem and is suitably modified to deal with power spectra.

In Section II we outline and discuss the axiomatic framework. In Section III we contrast the development herein with the axiomatic basis of Information Geometry and of alternative route to a spectral geometry based on prediction theory. In Section IV we present basic facts of the Monge-Katnorovic transportation problem which are then utilized in Section V, where we develop a suitable family of metrics satisfying the axioms of the sought spectral geometry.

Supported by Swedish Research Council, Royal Academy of Sciences, Göran Gustafsson Foundation, National Science Foundation and Air Force Office of Scientific Research; Johan Karlsson is with the Department of Mathematics, Division of Optimization and Systems Theory, Royal Institute of Technology, 100 44 Stockholm, Sweden, johan.karlsson@math.kth.se; Tryphon T. Georgiou and Mir Shahrouz Takyar are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA, {tryphon,shahrouz}@ece.umn.edu

II. Morphisms on power spectra

We consider power spectra of discrete-time stochastic processes. These are bounded positive measures on the interval $\mathbb{I} = (-\pi, \pi]$, and thus belong to

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

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 any suitable metric, result in small changes in any relevant statistical quantity.

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

$$R(\ell) := \mathcal{E}\{y(m)\overline{y(m-\ell)}\},$$

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 that are being considered in this paper, 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 $C_{\rm perio}((-\pi, \pi])$.

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 $y \cdot y_m$ is the circular convolution $d\nu = d\mu * d\mu_m$, i.e. $d\nu$ satisfies

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

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 their statistics. In such cases, additive noise 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. Thus, 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 pointwise 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 corresponding changes in statistics. More precisely, any topology induces a notion of convergence, and the question is whether this topology is compatible with the topology in the vector-space where 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 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 $\mathbf{G} \in C_{\rm perio}((-\pi, \pi])$.

III. Reflections and contrast with information geometry

The search for natural metrics between density functions can be traced back to several towering figures in the history of statistics, probability and information theory. A.N. Kolmogorov was "always interested in finding *information* distances" between probability distributions and, according to Chentsov [5, page 992] (ref. [1]), he independently arrived at and discussed the relevance of the Bhattacharyya [3] 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 $d\mu_0, d\mu_1$. Also according to Chentsov, A.N. Kolmogorov emphasized in his notes 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. Naturally, both suggestions reveal great intuition and foresight. The total variation admits the following interpretation (cf. [9]) that will turn out to be particularly relevant in our context. Assuming that $d\mu_0, d\mu_1$ are power spectra, the total variation represents the least "energy" of perturbations for $d\mu_0$ and $d\mu_1$ that render the two indistinguishable, i.e.,

$$\begin{aligned} d_{\rm TV}(d\mu_0, d\mu_1) \;\;=\;\; & \min\{\int d\nu_0 + \int d\nu_1 \;:\; d\nu_0,\, d\nu_1 \in \mathfrak{M}, \\ & \text{and } d\mu_0 + d\nu_0 = d\mu_1 + d\nu_1\} \end{aligned} \tag{2}$$

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 begun in the work of Fisher and Rao [12], [6], [2]. At the heart of the subject is the Fisher information metric on probability spaces and the closely related spherical Fisher-Bhattacharyya-Rao metric

$$d_{FBR}(d\mu_0, d\mu_1) := 2\arccos \int \sqrt{\mu_0(dx)\mu_1(dx)}. \tag{3}$$

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 states that stochastic maps are contractive with respect to the Fisher information metric and moreover, that this is the *unique* (up to constant multiple) Riemannian metric with this property [6]. 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

$$d_{\mathrm{KL}}(d\mu_0, d\mu_1) := \int \frac{d\mu_0}{d\mu_1} \log(\frac{d\mu_0}{d\mu_1}) d\mu_1 = \int \log(\frac{d\mu_0}{d\mu_1}) d\mu_0$$

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 [13]. 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 place on $\mathfrak{M}$. In particular, [7], [8] presented a metric for power spectra that quantifies the degradation of predictive error variance – in analogy with the latter argument that led to the Fisher metric. More precisely, a one-step optimal linear predictor for an underlying random process is obtained based on one power spectrum and then the 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

$$d_{\mathrm{predictive}}(d\mu_0, d\mu_1) := \sqrt{\int (\log \frac{d\mu_0}{d\mu_1})^2 d\theta - (\int \log \frac{d\mu_0}{d\mu_1} d\theta)^2}, \tag{4}$$

which effectively depends on the ratio of the corresponding spectral densities. A similar rationale can be based on smoothing instead of prediction (see [7], [8]), and this also leads to expressions that weigh in ratios of the corresponding spectral density functions.

A possible justification for such metrics, that 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 1: 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 satisfy 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

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

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

It is important to point out that none of the above is weak* continuous metrics. In particular, the metric in (4) is impervious to spectral lines as only the absolutely continuous part of the spectra play any role. Similarly, (2) and (3) cannot localize distributions either, based on their moments, because they also lack a needed weak* continuity. Thus, in this paper, we follow a line of reasoning analogous to the axiomatic framework of the Chentsov theorem, but for power spectra, requiring the metric to satisfy Axioms i)-iv).

IV. The Monge-Kantorovic problem

A natural class of metrics on measures are transport metrics based on the ideas of Monge and Kantorovic. The Monge-Kantorovic distance represents a cost of moving a nonnegative measure $d\mu_0 \in M(X)$ to another nonnegative measure $d\mu_1 \in M(X)$, 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. Every possible way of moving the measure $d\mu_0$ to $d\mu_1$ corresponds to a transference plan $\pi \in M(X \times X)$, 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

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

whenever $A, B \subset X$ are measurable. Such a plan exists only if the measures $d\mu_0$ and $d\mu_1$ has 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.

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

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

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

The optimal transportation problem admits a dual formulation, referred to as the Kantorovic duality (see [17]):

Theorem 2: Let c be a lower semi-continous (cost) function, let

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

and let

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

Then

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

Lemma 3: Let c be a lower semi-continous (cost) function with $c(x, x) = 0$ for $x \in X$. Then $A_{d\mu_\mathrm{a}}$ is contractive with respect to T_c.

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

$$\int_X \phi d\mu_0 + \psi d\mu_1 \ge \int_X \phi d\mu_0 + \psi d\mu_1 + (\phi + \psi)d\mu_\mathrm{a}.$$

■

Monge-Kantorovic distances are not metrics, in general, but they readily give rise to the so-called Wasserstein metrics. This is explained next.

Theorem 4: 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

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

is a metric and metrizes the weak* topology.

Proof: See [17], chapter 7. Note that since X is compact, the weak* topology on $\mathfrak{M}$ coincides with the weak topology. ■

V. Metrics based on transportation

The Monge-Kantorovic 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 nonnegative measures $d\mu_0$ and $d\mu_1$ on $\mathbb{I}$, we postulate that these are perturbations of the two measures $d\nu_0$ and $d\nu_1$, respectively, with 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

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) := \inf_{\nu_0(\mathbb{I})=\nu_1(\mathbb{I})} T_c(d\nu_0, d\nu_1) + \kappa \sum_{i=1}^{2} d_{\mathrm{TV}}(d\mu_i, d\nu_i), \tag{7}$$

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

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

where $(x)_{\mathrm{mod}2\pi}$ 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 5: Let $\kappa > 0$ and $c(x, y)$ defined as in (8), where $p \in (0, \infty)$. Then

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

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 6: Let c be a lower semi-continuous (cost) function, let

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

and let

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

Then

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

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 $M \ge \max\{\mu_i(\mathbb{I}) \,:\, i = 1, 2\}$. Accordingly, the (cost) function is modified as follows

$$\hat{c}(x, y) = \begin{cases} \min(c(x, y), 2\kappa) & \text{for} \quad x, y \in \mathbb{I}, \\ \kappa & \text{for} \quad x \in \mathbb{I}, y = \infty, \\ \kappa & \text{for} \quad x = \infty, y \in \mathbb{I}, \\ 0 & \text{for} \quad x = \infty, y = \infty. \end{cases} \tag{9}$$

Then $\tilde{T}_{c,\kappa}$ satisfies

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1), \tag{10}$$

and is, according to Theorem 2, equal to the supremum of

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

subject to

$$\begin{aligned} \phi(x)+\psi(y) &\le \hat{c}(x,y) && \text{for} \quad x,y\in\mathbb{I}, && (11)\\ \phi(x)+\psi(\infty) &\le \kappa && \text{for} \quad x\in\mathbb{I}, && (12)\\ \phi(\infty)+\psi(y) &\le \kappa && \text{for} \quad y\in\mathbb{I}, && (13)\\ \phi(\infty)+\psi(\infty) &\le 0. && && (14)\end{aligned}$$

The result now follows if we can show that there is no restriction to require $\phi(\infty)=\psi(\infty)=0$. This is since, $\Phi_{c,\kappa}$ is essentially identical to the set

$$\{(\phi,\psi) \,:\, (11)-(14) \text{ hold and } \phi(\infty)=\psi(\infty)=0\},$$

of the extensions of (ϕ,ψ) that now has support at ∞ as well.

To show this, let (ϕ,ψ) be an arbitrary pair of functions satisfying (11)-(14). 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

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

In both cases we have that $\hat{\mathcal{J}}(\phi,\psi)\le\hat{\mathcal{J}}(\hat{\phi},\hat{\psi})$ as well as that $(\hat{\phi},\hat{\psi})$ satisfies (11)-(14). In the second case, the constraint (13) 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 $\sup_{y\in\mathbb{I}}\psi(y)\le\kappa-\epsilon$.

In both cases, from an arbitrary pair (ϕ,ψ), we have constructed a pair $(\hat{\phi},\hat{\psi})$ for which the constraints (11)-(14) 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 $\phi(\infty)=\psi(\infty)=0$. ∎

Lemma 7: Let $c(x,y)$ be a function of $|x-y|$. Then for any $d\mu_{\rm m}\in\mathfrak{M}$ with $\int_{\mathbb{I}}d\mu_{\rm m}\le 1$, $M_{d\mu_{\rm m}}$ is contractive on $\mathfrak{M}$ with respect to $\tilde{T}_{c,\kappa}$.

Proof: Note that

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

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 in the functional $\mathcal{J}$.

Let $(\phi,\psi)\in\Phi_{c,\kappa}$, then we have that

$$\phi(x-\tau)+\psi(y-\tau)\le c(x-\tau,y-\tau),$$

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

$$\phi_m(x)+\psi_m(y)\le\min(c(x-\tau,y-\tau),2\kappa).$$

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

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

∎

Now we summarize the proof of the main theorem.

Proof: [Proof of Theorem 5]: From the formulation (10), $\tilde{T}_{c,\kappa}$ can be viewed as a transportation problem. Since the associated cost function $\hat{c}$ from (9) is of the form d^p, where d is a metric, Axiom i) and Axiom iv) follows from Theorem 4. From this formulation, Axiom ii) follows from Lemma 3. Finally Axiom iii) follows from Lemma 7. ∎

Remark 8: It is interesting to note that for the case $p=1$

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

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

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

VI. Concluding remarks

This work relates to a quantitative theory for spectral uncertainty. Our aim has been to identify natural notions of distance that allow localization of power spectra based on estimated statistics and, at the same time, share certain natural properties with regard to how noise affects distance between power spectra. We have presented an axiomatic framework that attempts to capture these intuitive notions and we have developed a family of metrics that satisfy the requirements.

While there are many possibilities for developing metrics with the required properties, we have chosen to base our approach on the concept of transportation. The reasons is that the resulting metrics have certain additional desirable properties which related to the deformations of spectra. 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 such geodesics, the corresponding formants are "matched" and the power transfers accordingly. It will be interesting to expand the set of axioms to include such desirable properties, which in turn may narrow down the possible choices of 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 (see [10]).

REFERENCES

[1] B.P. Adhikari and D.D. Joshi, "Distance, discrimination et resumé exaustif," *Publ. Inst. Univ. Paris*, **5**: 57-74, 1956.

[2] S. Amari and H. Nagaoka, **Methods of Information Geometry**, Translations of Mathematical Monographs, AMS, Oxford University Press, 2000.

[3] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, **35**: 99-109, 1943.

[4] E.F. Beckenbach and R. Bellman, **Inequalities**, Springer-Verlag, Berlin-Heidelberg, 198 pages,1965.

[5] N.N. Chentsov, "The unfathomable influence of Kolmogorov," *The Annals of Statistics*, **18(3)**: 987-998, 1990.

[6] N.N. Chentsov, **Statistical Decision Rules and Optimal Inference**, Nauka, Moscow. English translation, Providence 1982.

[7] T.T. Georgiou, "Distances between power spectral densities," *IEEE Trans. on Signal Processing*, **55(8)**: 3993-4003, August 2007.

[8] T.T. Georgiou, "An intrinsic metric for power spectral density functions," *IEEE Signal Processing Letters,* **14(8):** 561-563, August 2007.

[9] T.T. Georgiou, "Distances Between Time-Series and Their Autocorrelation Statistics," **Modeling, Estimation and Control**, Festschrift in Honor of Giorgio Picci, Eds. A. Chiuso, A. Ferrante, and S. Pinzoni, Springer-Verlag, Berlin, pp. 113-122, 2007.

[10] T.T. Georgiou, C.R. Shankwitz, and M.C. Smith, "Identification of linear systems: A graph point of view," American Control Conf., 1982.

[11] U. Grenander and G. Szegö, **Toeplitz Forms and their Applications**, Chelsea, 1958.

[12] R.E. Kass, "The geometry of asymptotic inference," *Statistical Science*, **4(3)**): 188-234, 1989.

[13] S. Kullback, **Information Theory and Statistcs**, Dover, 1997.

[14] J. Karlsson and T.T. Georgiou, "Signal analysis, moment problems & uncertainty measures," Proceedings of the IEEE Intern. Conf. on Decision and Control, pp. 5710- 5715, December 2005.

[15] S. Rachev and L Rüschendorf, **Mass transportation Problems, Volume 1: Theory**, Springer, 1998.

[16] P. Stoica and R. Moses, **Introduction to Spectral Analysis**, Prentice Hall, 2005.

[17] C. Villani, **Topics in Optimal Transportation**, Graduate studies in Mathematics vol 58, AMS, 2003.