Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008



# Transport metrics for power spectra

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

*Abstract*— **We present a family of metrics for power spectra based on the Monge-Kantorivic transportation distances. These metrics are constructed so that distances reduce with additive and multiplicative noise, reflecting the intuition that noise typically reduces our ability to discriminate spectra. In addition, perturbations measured in these metrics are continuous with respect to the statistics of the underlying time series. A general framework for constructing such metrics is put forth and these are contrasted with an earlier Riemannian metric which is based on prediction theory and the relevant geometry of the underlying time-series.**

## I. INTRODUCTION

Our goal in this paper is to motivate and develop metrics between power spectra that have a number of natural properties. More specifically, since noise impedes our ability to discriminate it is natural to seek metrics for which the distance between two power spectra does not increase when the same amount of noise power is added to both. Similarly, multiplicative noise of unit variance must also not increase their distance. Finally, since power spectra are invariably estimated using second-order statistics, it is important that statistics and spectra relate in a continuous manner. Naturally, such a property depends on the metric that we use for assessing perturbation from a nominal power spectrum. Thus, we are interested in metrics for which any statistic is continuous with respect to spectral uncertainty. The usefulness of such a property, known as weak$^*$ continuity, is further explored in [14]. The present paper is based on [9] where we refer the reader for detailed proofs of the various propositions.

Thus, below, we first motivate and discuss a set of axioms that from a certain viewpoint are natural for characterizing spectral metrics (Section II). We then draw analogies and contrasts with the setting of Information Geometry which deals with probability distributions instead of spectra (Section III). In Section III we also outline an alternative route to a spectral geometry based on prediction theory which was introduced in [6], [7]. Returning to the type of metric properties that are sought in Section II, weak$^*$ continuity of power spectral distances readily suggests Monge-Kantorovich transportation between measures as a suitable framework for constructing such metrics. Thus, in Section IV, we introduce the Monge-Kantorovich transportation problem and review certain basic results. These results are then utilized in Section V where we develop a suitable family of metrics satisfying the axioms of the sought spectral geometry. We conclude in Section VI with an example of power spectra having spectral lines which highlights differences between our metrics, the well-known Itakura-Saito distance, and the prediction-related metric introduced in [6], [7].

## II. TRANSFORMATIONS 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 transformations 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 $2\pi$. For future reference we denote the set of such functions by $C_{\text{perio}}((-\pi, \pi])$.

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 $2\pi$.

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 transformations on $\mathfrak{M}$ that transform power spectra accordingly. Additive and multiplicative noise transformations 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 transformations. 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. 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 [4, 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. [8]) 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], [5], [2]. At the heart of the subject is the Fisher information metric on probability spaces and the closely related spherical Fisher-Bhattacharyya-Rao metric

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

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 [5]. 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 $d_{\mathrm{KL}}$, 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 $d_{\mathrm{TV}}$ 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 $\mathfrak{M}$. In particular, [6], [7] 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 [6], [7]), 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 transformations 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, $\mu$ 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 $\delta(\cdot,\cdot)$.

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-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 $d\mu_0 \in M(X)$ 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 Kantorovich duality (see [15]):

*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 $A_{d\mu_{\rm a}}$ is contractive with respect to $T_c$.

Monge-Kantorovich 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 on the subspace of $\mathfrak{M}(X)$ with fixed mass and metrizes 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 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

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

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

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

where $(x)_{\rm mod2\pi}$ 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

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

*Remark 7:* 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 6 is amenable to numerical implementation. Indeed, (9) is a linear optimization problem which can be computed using standard methods.

*Lemma 8:* 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 $\tilde{T}_{c,\kappa}$.

Now we sketch the proof of Theorem 5. The complete proof can be found in [9]. In view of Theorem 6, $\tilde{T}_{c,\kappa}$ can be viewed as the cost of a transportation problem with cost function of the form $d^p$, where $d$ is a metric. Therefore, Axiom i) and Axiom iv) follows from Theorem 4. From this formulation, Axiom ii) follows from Lemma 3. Finally Axiom iii) follows from Lemma 8.

*Remark 9:* It is interesting to note that for the case $p = 1$

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

where $\|g\|_L = \sup \frac{|g(x)-g(y)|}{|x-y|}$ denotes 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. Example

Next, we present an example that highlights the relevance of the proposed metrics in spectral analysis. The example compares how different distance measures perform on spectra which contain spectral lines. The distance measures we consider, besides the transportation distance (here $\delta_{1,1}$), are the prediction metric and the Itakura-Saito distance.

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, $\theta$ is taken as a constant, whereas $\phi$ is assumed random, independent of $w_k$, and uniformly distributed on $(-\pi,\pi]$. Figure 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 (periodogram). 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 [7], and iii) the Itakura-Saito distance (see e.g., [10]). Comparison of these values reveals that only the transportation-based metric can reliably distinguish between spectral lines.

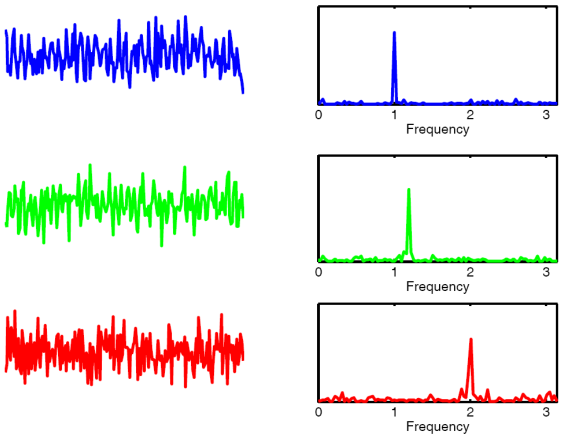


Fig. 1. Stochastic process $y_k$ in time and frequency domain for $\theta = 1, 1.2,$ and 2

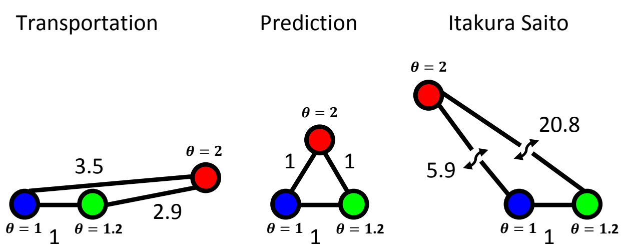


Fig. 2. Relative distances between line spectra

The schematic in Figure 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

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: $d_{\rm pr}$ | 1.8428 | 1.8390 | 1.8517 |
| Itakura Saito: $d_{IS}$ | 22.7279 | 472.6690 | 134.1707 |

$\{1, 1.2\}$ 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 their true values. On the other hand, this is not the case for either of the other two distances.

## VII. Concluding Remarks

Our aim has been to identify natural notions of distance and to develop a quantitative theory for spectral uncertainty that allows 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 presented an axiomatic framework that attempts to capture these intuitive notions and we developed a family of metrics that satisfy the stated requirements.

While there are many possible metrics with the required properties, we have chosen to base our approach on the concept of transportation. The resulting metrics appear to have certain additional qualities. More specifically, from experience, it appears that geodesics (in e.g., the Wasserstein 2-metric) match "formants" between the two end points in a more natural way. Such a property may be useful in speech processing (cf., see [11]). It is interesting to speculate on what type of additional requirements/axioms may lead to a unique metric. Finally, we remark that there is a need for analogous metrics for comparing multivariable spectra.

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

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

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

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

[8] 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.

[9] T.T. Georgiou, J. Karlsson, M.S. Takyar, "Metrics for power spectra: an axiomatic approach", *IEEE Trans. on Signal Processing*, to appear.

[10] R. Gray, A. Buzo, A. Gray, and Matsuyama, "Distortion measures for speech processing," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. 28, no. 4, Aug. 1980.

[11] 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, (Blondel V., Boyd S., Kimura H. eds.), Festschrift on the occasion of the 60-th birthday of M. Vidyasagar, vol. 371, Springer Verlag 2008.

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