

Distances and Riemannian Metrics for Spectral Density Functions

Tryphon T. Georgiou, *Fellow, IEEE*

Abstract—We introduce a differential-geometric structure for spectral density functions of discrete-time random processes. This is quite analogous to the Riemannian structure of information geometry, which is used to study perturbations of probability density functions, and which is based on the Fisher information metric. Herein, we introduce an analogous Riemannian metric, which we motivate with a problem in prediction theory. It turns out that this problem also provides a prediction theoretic interpretation to the Itakura distortion measure, which relates to our metric. Geodesics and geodesic distances are characterized in closed form and, hence, the geodesic distance between two spectral density functions provides an explicit, intrinsic (pseudo)metric on the cone of density functions. Certain other distortion measures that involve generalized means of spectral density functions are shown to lead to the same Riemannian metric. Finally, an alternative Riemannian metric is introduced, which is motivated by an analogous problem involving smoothing instead of prediction.

Index Terms—Differential structure, distortion measures, information geometry, metrics, spectral density functions.

I. INTRODUCTION

DUE to the centrality of spectral analysis in a wide range of scientific disciplines, there has been a variety of viewpoints regarding how to quantify distances between spectral density functions. Besides the obvious ones, which are based on norms, inherited by ambient function-spaces L_2 , L_1 , etc., there has been a plethora of alternatives that attempt to acknowledge the structure of power spectral density functions as a positive cone. The most well known are the Kullback–Leibler divergence, which originates in hypothesis testing, Bayes’ estimation and coding, the Itakura–Saito distance, which originates in speech analysis—both belonging to the Bregman class [6], [15], the Hellinger [14] and Bhattacharyya distances [4], and the Ali–Silvey class of divergences [3]. Their origin can be traced either to a probabilistic rationale (as in the case of the Kullback–Leibler divergence) or to some ad hoc mathematical construct designed to seek distance measures with desirable properties (as in the case of Bregman and Ali–Silvey classes). Often, symmetry and the triangular inequality are dispensed of, in which case the term “distortion measure” is used [15]. Our intent in this paper is not to compare the vast number of possible distortion measures, but instead, to motivate and

develop a natural metric and a geometry for the cone of spectral densities.

A model for our approach is that of information geometry. In it, one may begin with the Kullback–Leibler divergence as quantifying degradation of coding efficiency in terms of average word length, when a code is designed based on one probability distribution and then applied to another. The degradation of efficiency quantifies distance between two probability distributions. Differential geometry is then introduced when one considers varying the probability distribution directly or, possibly, in a suitable parameter space. The natural Riemannian metric, which follows from the Kullback–Leibler divergence is the Fisher information metric g_{Fisher} [1], [2]. The Kullback–Leibler divergence is one of several possible “divergences” that give rise to the same differential structure. Interestingly, geodesics and geodesic distances are computable in closed form. We follow a similar path with remarkably analogous, but different, results. These similarities are further highlighted in Section VI.

Our starting point is a prediction problem for random process. We select an optimal predictive filter for an underlying random process based on the assumption that the process has $f_1(\theta)$ as its spectral density. We then evaluate the performance of such a filter against a second random process with spectral density $f_2(\theta)$. Naturally, the degradation of predictive error variance can be considered as quantifying the mismatch between the two spectral density functions. The degradation (actual error variance divided by minimal error variance) turns out to equal to the ratio of the arithmetic over the geometric mean of the fraction of the two power spectra. Interestingly, the logarithm of the degradation coincides with the Itakura distortion measure [15, p. 371] between the two spectral densities. Herein, the degradation of predictive error variance and the Itakura distance play a role similar to the one played by the Kullback–Leibler distance in information geometry.

The degradation of predictive error variance between infinitesimal perturbations of power spectra gives rise to a Riemannian metric on the manifold of spectral density functions. This is analogous to the Fisher information metric. Using variational calculus, we characterize geodesics. The geodesic connecting two spectral density functions turns out to constitute an exponential family. Alternatively, geodesics are more clearly understood if we use the logarithm to map spectral density functions into a linear space. In this linear space, essentially, geodesics are *straight lines*. The geodesic distance can be readily computed in closed form and, thus, it represents an intrinsic (pseudo)metric. The metric is expressed as the variance of the logarithm of the ratio of the two spectral density functions. It is a pseudometric because it does not detect

Manuscript received July 29, 2006; revised January 2, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Peter Handel. This research has been supported by the National Science Foundation and the Air Force Office of Scientific Research.

The author is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (tryphon@ece.umn.edu).

Digital Object Identifier 10.1109/TSP.2007.896119

scaling by multiplicative constants. Other than that, being a geodesic distance, it is an intrinsic metric and satisfies all other metric properties.

An alternative concept of distance can be built based on the degradation of the variance of smoothing filters. As in the case of our prediction problem, a smoothing filter is now designed based on one density function and then applied to another. The degradation of smoothing error variance quantifies the mismatch between the two spectral density functions. A Riemannian metric can be defined from this as well in a similar manner.

The paper is structured as follows. Section II presents notation and needed background material on linear prediction and smoothing. In Section III, we compute the degradation of predictive error variance. Its logarithm turns out to coincide with the Itakura distance measure. Section IV introduces the relevant Riemannian structure, characterizes geodesics, and computes geodesic distances. Section V outlines an analogous Riemannian metric, which is based on the degradation of smoothing-error variance. In Section VI, we overview certain key concepts from information geometry and draw parallels with our development regarding power spectra. Section VII contains some final remarks.

II. PRELIMINARIES ON LEAST-VARIANCE PREDICTION AND SMOOTHING

Consider a scalar zero-mean stationary random process $\{u_k, k \in \mathbb{Z}\}$ and denote by $d\mu(\theta)$ its power spectrum. We are interested in quadratic optimization problems with respect to the usual inner product

$$\left\langle \sum_k a_k u_k, \sum_k b_k u_k \right\rangle := \mathcal{E} \left\{ \left(\sum_k a_k u_k \right) \left(\sum_k b_k u_k \right)^* \right\}$$

where \mathcal{E} denotes expectation and “*” denotes complex conjugation. The closure of $\text{span}\{u_k, k \in \mathbb{Z}\}$, which we denote by \mathcal{U} , can be identified with the space $L_{2,d\mu}[-\pi, \pi]$ of functions, which are square integrable with respect to the spectral measure $d\mu(\theta)$ with inner product

$$\langle a, b \rangle_{d\mu} := \frac{1}{2\pi} \int_{-\pi}^{\pi} a(\theta) (b(\theta))^* d\mu(\theta)$$

where $a(\theta) = \sum_k a_k e^{jk\theta}$ and $b(\theta) = \sum_k b_k e^{jk\theta}$. Further, the correspondence

$$\mathcal{U} \rightarrow L_{2,d\mu} : \sum_k a_k u_k \mapsto \sum_k a_k e^{jk\theta}$$

is a Hilbert space isomorphism (see [17]), and thus, least-variance approximation problems can be equivalently expressed in $L_{2,d\mu}$.

In particular, the variance of the *one-step-ahead prediction* error $u_0 - \hat{u}_{0|\text{past}}$ for the *predictor*

$$\hat{u}_{0|\text{past}} = \sum_{k>0} \alpha_k u_{-k}$$

is

$$\mathcal{E} \{ |u_0 - \hat{u}_{0|\text{past}}|^2 \} = \left\| 1 - \sum_{k>0} \alpha_k e^{jk\theta} \right\|_{d\mu}^2. \quad (1)$$

Similarly, the variance of the error of the *smoothing* filter

$$\hat{u}_{0|\text{past \& future}} := \sum_{k \neq 0} \beta_k u_{-k} \quad (2)$$

is simply

$$\mathcal{E} \{ |u_0 - \hat{u}_{0|\text{past \& future}}|^2 \} = \left\| 1 - \sum_{k \neq 0} \beta_k e^{jk\theta} \right\|_{d\mu}^2. \quad (3)$$

In general, the power spectrum $d\mu$ is a bounded nonnegative measure on $[-\pi, \pi)$ and admits a decomposition $d\mu = d\mu_s + fd\theta$ with $d\mu_s$ a singular measure and $fd\theta$ the absolutely continuous part of $d\mu$ (with respect to the Lebesgue measure). In general, the singular part has no effect on the minimal variance of the error, and the corresponding component of u_k can be estimated with arbitrary accuracy using any “one-sided” infinite past. The variance of the optimal one-step-ahead prediction error depends only on the absolutely continuous part of the power spectrum and is given in terms by the celebrated Szegő–Kolmogorov formula stated below (see [20] and also [12, p. 183], [13], [19], and [21, Ch. 6]).

Theorem 1: With $du = d\mu_s + fd\theta$ as above

$$\inf_{\alpha} \left\| 1 - \sum_{k>0} \alpha_k e^{jk\theta} \right\|_{d\mu}^2 = \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\theta) d\theta \right\}$$

when $\log f \in L_1[-\pi, \pi)$, and zero otherwise.

In case $\log f \in L_1[-\pi, \pi)$, the prediction-error variance is nonzero and the random process is nondeterministic in the sense of Kolmogorov. In this case, it can be shown that

$$f(\theta) = \frac{gf}{|a_f(e^{j\theta})|^2}$$

where $a_f(z)$ is an outer function in the Hardy space $H_2(\mathbb{D})$ with $a_f(0) = 1$, i.e.,

$$a_f(z) = 1 + a_{1,f}z + a_{2,f}z^2 + \dots \quad (4)$$

is analytic in the unit disk $\mathbb{D} := \{z : |z| < 1\}$, and its radial limits are square integrable (see [18]). Then, the linear combination

$$\hat{u}_{0|\text{past}} := \sum_{k>0} (-a_{k,f}) u_{-k} \quad (5)$$

serves as the optimal predictor of u_0 based on past observations and the least variance of the optimal prediction error becomes

$$\begin{aligned} \mathcal{E} \left\{ \left| u_0 - \sum_{k>0} (-a_{k,f}) u_{-k} \right|^2 \right\} &= \lim_{r \nearrow 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |a(re^{j\theta})|^2 f(\theta) d\theta \\ &= \exp \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f(\theta)) d\theta \right) \\ &=: g_f. \end{aligned}$$

Analogous expressions exist for the optimal smoothing error and the corresponding smoothing filter, which uses both past and future values of u_k . It is quite interesting, and rather straightforward, that while the variance of the optimal one-step-ahead prediction error is the *geometric mean* of the spectral density function, the variance of the error, when a smoothing filter utilizes both past and future, turns out to be the *harmonic mean* of the spectral density function. This is stated below—for a proof, see [11].

Theorem 2: With $d\mu = d\mu_s + f d\theta$ as above

$$\inf_{\beta} \left\| 1 - \sum_{k \neq 0} \beta_k e^{jk\theta} \right\|_{d\mu}^2 = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)^{-1} d\theta \right)^{-1} =: h_f \tag{6}$$

when $f^{-1} \in L_1[-\pi, \pi]$, and zero otherwise.

In case $f^{-1} \in L_1[-\pi, \pi]$ the variance of the optimal smoothing error is nonzero and the random process is non-deterministic in the sense that past and future specify the present, which can be estimated with zero variance. In this case (cf. [11])

$$\begin{aligned} b_f(\theta) &:= \dots + b_{-2,f} e^{-2j\theta} + b_{-1,f} e^{-j\theta} + 1 \\ &\quad + b_{1,f} e^{j\theta} + b_{2,f} e^{2j\theta} + \dots \\ &= h_f f(\theta)^{-1} \end{aligned} \tag{7}$$

is the image of the optimal smoothing error $u_0 - \sum_{k \neq 0} (-b_k) u_k$ under the Kolmogorov map, and

$$\begin{aligned} \mathcal{E} \left\{ \left| u_0 - \sum_{k \neq 0} (-b_k) u_k \right|^2 \right\} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |b_f(\theta)|^2 f(\theta) d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} h_f^2 (f(\theta))^{-2} f(\theta) d\theta \\ &= h_f. \end{aligned}$$

III. DEGRADATION OF THE PREDICTION-ERROR VARIANCE

We now consider two distinct spectral density functions f_1 and f_2 and postulate a situation where filtering of an underlying random process is attempted based on the incorrect choice between these two alternatives. The variance is then compared

with the least possible variance, which is achieved when the correct choice is made (i.e., when the predictor is optimal for the spectral density against which it is being evaluated). The degradation of performance is quantified by how much the ratio of the two prediction-error variances exceeds the identity.

This ratio may serve as a measure of mismatch between the two spectral densities (the one that was used to design the predictor and the one against which it is being evaluated). The resulting mismatch turns out to be *scale-invariant*—i.e., the expression is homogeneous. Hence, as a measure of distance, it actually quantifies distance between the positive rays that the two spectral density functions define, and thus, it quantifies distance between the respective “shapes.” It turns out that this distance is log-convex on logarithmic intervals and has a number of distancelike properties, short of being a metric.

Proposition 3: Consider two second-order stationary non-deterministic random processes $\{u_k^{(i)}, k \in \mathbb{Z}\}$ having power spectral densities f_i , for $i \in \{1, 2\}$, respectively. The degradation of prediction-error variance defined as

$$\rho(f_1, f_2) := \frac{\mathcal{E} \left\{ \left| u_0^{(1)} - \sum_{k=1}^{\infty} a_{k,f_2} u_{-k}^{(1)} \right|^2 \right\}}{\mathcal{E} \left\{ \left| u_0^{(1)} - \sum_{k=1}^{\infty} a_{k,f_1} u_{-k}^{(1)} \right|^2 \right\}} \tag{8}$$

where a_{k,f_i} represent the coefficients of optimal linear predictors defined in (4) and (5), is

$$\rho(f_1, f_2) = \frac{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta) d\theta}{f_2(\theta) 2\pi} \right)}{\exp \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(\frac{f_1(\theta)}{f_2(\theta)} \right) \frac{d\theta}{2\pi} \right)}. \tag{9}$$

Proof: The proof is a trivial application of classical prediction theory summarized in the previous section. Indeed, if the predictor is based on f_2 whereas the underlying process has f_1 as its spectral density, then the variance of the prediction error is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |a_{f_2}(e^{j\theta})|^2 f_1(\theta) d\theta = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta \right) g_{f_2}.$$

If we divide this variance by the optimal value g_{f_1} we obtain

$$\rho(f_1, f_2) = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta \right) \frac{g_{f_2}}{g_{f_1}}. \quad \blacksquare$$

The expression for $\rho(f_1, f_2)$ in (9) is the ratio of the *arithmetic mean* over the *geometric mean* of the fraction f_1/f_2 of the two spectral density functions. It is not symmetric in the two arguments. It can be readily seen, as a consequence of Jensen’s inequality, that $\rho(f_1, f_2) > 1$ unless f_1/f_2 is constant, in which case $\rho(f_1, f_2) = 1$.

Since $\rho(f_1, f_2) \geq 1$, the expressions

$$\rho(f_1, f_2) - 1 =: \gamma(f_1, f_2) \tag{10a}$$

and

$$\begin{aligned} \log(\rho(f_1, f_2)) &= \log\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta\right) \\ &\quad - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(\frac{f_1(\theta)}{f_2(\theta)}\right) d\theta \\ &=: \delta(f_1, f_2) \end{aligned} \quad (10b)$$

take values in $[0, \infty]$. Their value is zero only when the ratio f_1/f_2 is constant. Thus, either expression may be used to quantify the distance between f_1 and f_2 .

Remark 4: Interestingly, $\delta(f_1, f_2)$ was arrived at earlier by Itakura (cf. [15, p. 371, Sec. C], [16]) following a different rationale—it was introduced as a “gain optimized” distortion for use in speech recognition systems, and as a lower bound to the Itakura–Saito distortion. \square

Viewing $\gamma(f_1, f_2)$ and $\delta(f_1, f_2)$ as quantifying the slackness in a Jensen-type inequality suggests a variety of other analogous expressions based on generalized means that may be used to quantify flatness of $\Lambda := f_1/f_2$. More specifically, if

$$M_r(\Lambda) := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \Lambda(\theta)^r d\theta\right)^{1/r}$$

denotes the r th generalized mean of function $\Lambda(\theta)$, then

$$M_r(\Lambda) \leq M_s(\Lambda) \quad \text{for any } -\infty \leq s < r \leq \infty$$

(see [5]). Hence

$$\gamma_{r,s}(f_1, f_2) := \frac{M_r(f_1/f_2)}{M_s(f_1/f_2)} - 1 \geq 0 \quad \text{and} \quad (11a)$$

$$\begin{aligned} \delta_{r,s}(f_1, f_2) &:= \log(M_r(f_1/f_2)) \\ &\quad - \log(M_s(f_1/f_2)) \geq 0 \end{aligned} \quad (11b)$$

with a value that depends on how “far” Λ is from being constant. Note that

$$\gamma(f_1, f_2) = \gamma_{1,0}(f_1, f_2) \quad (12a)$$

$$\delta(f_1, f_2) = \delta_{1,0}(f_1, f_2) \quad (12b)$$

since

$$M_0(\Lambda) := \lim_{r \rightarrow 0} M_r(\Lambda) = e^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\Lambda(\theta)) d\theta} \quad (13)$$

is in fact the geometric mean of Λ (see, e.g., [5, p. 23]).

Of particular interest is the following possible symmetrized version of the Itakura distance $\delta_{1,0}$:

$$\begin{aligned} \delta_{1,0}(f_1, f_2) + \delta_{1,0}(f_2, f_1) &= \log\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta\right) \\ &\quad + \log\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_2(\theta)}{f_1(\theta)} d\theta\right) \\ &=: \delta_{1,-1}(f_1, f_2) \end{aligned}$$

which is in fact the logarithm of the ratio of the arithmetic mean over the harmonic mean of the “likelihood” ratio f_1/f_2 .

IV. RIEMANNIAN STRUCTURE ON POWER SPECTRA

Infinitesimal perturbations about a given power spectral density function, when measured by any of $\delta(\cdot, \cdot)$, or $\delta_{r,s}(\cdot, \cdot)$ (with $r > s$), give rise to the same nonnegative definite quadratic form. The quadratic form varies continuously with the choice of spectral density about which perturbations are being considered, and therefore, it constitutes a Riemannian metric. As it follows from the previous section, this Riemannian metric is quite natural from a prediction-theory point of view as it measures the degradation of predictive error variance for small perturbations. For large perturbations, the metric can be integrated along paths that represent the deformation of one spectral density into another. When such deformations/paths are being selected so that the overall distance between the end points (power spectra) is minimal, they constitute geodesics. The geodesic distance between two power spectra defines a natural metric.

The geometry induced by a Riemannian metric is interesting in its own right. However, in general, the computation of geodesics and geodesic distances is a nontrivial task. Thus, it is quite unexpected that, for our Riemannian metric, as we will see in this section, geodesics and geodesic distances can in fact be computed in closed form (by expressions that involve the end power spectra). It turns out that the map

$$\log : f \mapsto \log(f)$$

which takes the cone of power spectral densities into a linear space is quite natural with respect to our Riemannian metric in that geodesics map into straight lines. This is the reason for the simplicity of the induced Riemannian geometry and provides a further justification for the common practice of drawing insight from logarithmic plots of power spectra.

Throughout, we assume that all functions are smooth enough so that the indicated integrals exist. This, in particular, can be ensured if all spectral density functions are bounded and have bounded inverses as well as bounded derivatives. Weaker conditions are clearly possible. For the purposes of this paper, we define

$$\mathcal{F} := \left\{ f : f \text{ differentiable on } [-\pi, \pi], \right. \\ \left. \text{with } f(\theta) > 0, \text{ and } f, \frac{df}{d\theta}, f^{-1} \text{ bounded} \right\}$$

and recall the definitions of the k th norm and the supnorm

$$\|f\|_k := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^k d\theta\right)^{1/k}$$

and the supnorm $\|f\| := \sup\{|f(\theta)| : \theta \in [-\pi, \pi]\}$. The natural tangent space of \mathcal{F} where perturbations may belong is $\mathcal{D} := \{\Delta : \Delta, (d\Delta/d\theta) \text{ bounded}\}$.

Proposition 5: Let $f, f + \Delta \in \mathcal{F}$, where Δ such that $|\Delta/f| < 1$, and $r, s \in \mathbb{R}$ with $r > s$. Then

$$\begin{aligned} \gamma_{r,s}(f + \Delta, f) &= \delta_{r,s}(f + \Delta, f) + O(\|\Delta/f\|^3) \\ &= \left(\frac{r-s}{2}\right) \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)}\right)^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta\right)^2 \right) + O(\|\Delta/f\|^3). \end{aligned} \tag{14}$$

Here, $O(\|\Delta/f\|^3)$ indicates terms of order 3 or higher.

The quadratic form that appears in (15) may be thought of as the “variance” of the integrand Δ/f . Similar expressions will appear often; hence, we introduce the notation

$$\|\varphi\|_{\text{var}}^2 := \frac{1}{2\pi} \int_{-\pi}^{\pi} (\varphi(\theta))^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\theta) d\theta\right)^2 \tag{16}$$

and note that $\|\varphi + c\|_{\text{var}} = \|\varphi\|_{\text{var}}$ for any c that is independent of θ .

Proof: If $\Lambda(\theta) = 1 + \epsilon(\theta)$ represents a small perturbation of the identity for $\theta \in [-\pi, \pi]$, then

$$\begin{aligned} M_r(\Lambda) &= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + \epsilon(\theta))^r d\theta\right)^{1/r} \\ &= \left(1 + r \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta) d\theta + \frac{r(r-1)}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta)^2 d\theta + O(\|\epsilon(\theta)\|^3)\right)^{1/r} \\ &= 1 + \frac{1}{r} \left(r \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta) d\theta + \frac{r(r-1)}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta)^2 d\theta\right) \\ &\quad + \frac{1}{2} \frac{1-r}{r^2} \left(r \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta) d\theta\right)^2 + O(\|\epsilon(\theta)\|^3) \\ &= 1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta) d\theta + \frac{r-1}{2} \|\epsilon\|_{\text{var}}^2 + O(\|\epsilon\|^3). \end{aligned}$$

When $r = 0$, the same holds true. Indeed, using the series $\log(1 + \epsilon) = \epsilon - (1/2)\epsilon^2 + (1/3)\epsilon^3 - \dots$ and in view of (13)

$$\begin{aligned} M_0(\Lambda) &= \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\epsilon(\theta) - \frac{1}{2}\epsilon(\theta)^2 + \dots\right) d\theta \right\} \\ &= 1 + \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\epsilon(\theta) d\theta - \frac{1}{2}\epsilon(\theta)^2 + \dots\right) d\theta\right) \\ &\quad + \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\epsilon(\theta) d\theta - \frac{1}{2}\epsilon(\theta)^2 + \dots\right) d\theta\right)^2 \\ &\quad + O(\|\epsilon\|^3) \end{aligned}$$

$$= 1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \epsilon(\theta) d\theta - \frac{1}{2} \|\epsilon\|_{\text{var}}^2 + O(\|\epsilon\|^3).$$

Using the above, we observe that

$$\begin{aligned} \gamma_{r,s}(f + \Delta, f) &= \frac{M_r(1 + \Delta/f) - M_s(1 + \Delta/f)}{M_s(1 + \Delta/f)} \\ &= \frac{\frac{r-s}{2} \|\Delta/f\|_{\text{var}}^2 + O(\|\Delta/f\|^3)}{1 + O(\|\epsilon\|)} \\ &= \frac{r-s}{2} \|\Delta/f\|_{\text{var}}^2 + O(\|\Delta/f\|^3) \end{aligned}$$

as claimed. Similarly, since

$$\begin{aligned} \log \left(M_r \left(\frac{f + \Delta}{f} \right) \right) - \log \left(M_s \left(\frac{f + \Delta}{f} \right) \right) &= \frac{M_r \left(\frac{f + \Delta}{f} \right)}{M_s \left(\frac{f + \Delta}{f} \right)} - 1 + O \left(\left(\frac{M_r \left(\frac{f + \Delta}{f} \right)}{M_s \left(\frac{f + \Delta}{f} \right)} - 1 \right)^2 \right) \end{aligned}$$

it follows that $\delta_{r,s}(f + \Delta, f)$ is identical to $\gamma(f + \Delta, f)$ up to third-order terms. We note that when dealing with higher order terms, we use the fact that $\|\Delta/f\|_k \leq \|\Delta/f\|$ for all k , since $|\Delta/f| < 1$. ■

Proposition 5 suggests the following natural Riemannian metric on the cone \mathcal{F} of spectral density functions.

Definition 6: For any $f \in \mathcal{F}$, we define the Riemannian metric

$$g_f : \mathcal{D} \rightarrow \mathbb{R}_+ : \Delta \mapsto g_f(\Delta) := \left\| \frac{\Delta}{f} \right\|_{\text{var}}^2. \tag{17}$$

This is indeed a Riemannian metric as it represents a continuously varying non-negative definite quadratic form for any $f \in \mathcal{F}$. The fact that the quadratic form $g_f(\Delta)$ is non-negative is a direct consequence of the Cauchy–Schwartz inequality. Next, we characterize the geodesics of our Riemannian metric.

Thus, we consider continuous paths $f_\tau \in \mathcal{F}$, $\tau \in [0, 1]$, connecting two given spectral density functions f_0 and f_1 . A path is to be thought as a continuous deformation of f_0 into f_1 . Therefore, f_τ is a function of two arguments, the path parameter τ and the frequency θ —hence, we write $f_\tau(\theta)$ when the frequency θ needs to be specified. We denote by \dot{f}_τ the partial derivative of f_τ with respect to τ , i.e.,

$$\dot{f}_\tau(\theta) := \frac{\partial f_\tau(\theta)}{\partial \tau}.$$

The partial derivative for a particular value for τ is, of course, a function of θ and represents the rate that f_τ is deforming. The length traversed as τ varies from 0 to 1 is

$$\begin{aligned} \ell(f_\tau : 0, 1) &= \int_0^1 \sqrt{g_{f_\tau}(\dot{f}_\tau d\tau)} = \int_0^1 \left\| \frac{\dot{f}_\tau(\theta)}{f_\tau(\theta)} \right\|_{\text{var}} d\tau \\ &= \int_0^1 \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\dot{f}_\tau(\theta)}{f_\tau(\theta)}\right)^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\dot{f}_\tau(\theta)}{f_\tau(\theta)} d\theta\right)^2} d\tau. \end{aligned} \tag{18}$$

Paths of minimal length are called geodesics.

Geodesics connecting two power spectra are not necessarily unique. This is indeed the case for our Riemannian metric. In fact, if f_τ is one particular geodesic, then for any $\alpha(\tau)$ that is a continuous function of τ , independent of θ , and satisfies $\alpha(0) = \alpha(1) = 1$, $\alpha(\tau)f_\tau$ is also a geodesic. To see this, it suffices to note that the integrand in (18) is independent of $\alpha(\tau)$ since

$$\left\| \frac{\frac{\partial}{\partial \tau}(\alpha f_\tau)}{\alpha f_\tau} \right\|_{\text{var}} = \left\| \left(\frac{\dot{f}_\tau}{f_\tau} + \frac{\dot{\alpha}}{\alpha} \right) \right\|_{\text{var}} = \left\| \frac{\dot{f}_\tau}{f_\tau} \right\|_{\text{var}}$$

where the last equality follows from the fact that $\dot{\alpha}/\alpha$ is independent of θ . This can also be traced to the fact that the degradation of predictive-error variance $\rho(f_1, f_2)$ in (9), on which $g_f(\cdot)$ is based, is unaffected by scaling of either of its arguments.

For any path f_τ between f_0 and f_1 , a rescaling of the indexing leaves the length traversed unaffected (and only affects the “velocity”). Thus, if f_τ is a geodesic, then $f_{\beta(\tau)}$ is also a geodesic for any continuous, monotonically nondecreasing function β such that $\beta(0) = 0$ and $\beta(1) = 1$. It turns out that, modulo the above two possibilities, the geodesic connecting two power spectra is essentially unique and can be written in the form of a logarithmic interval (otherwise, exponential family)

$$\begin{aligned} f_\tau &= e^{(1-\tau)\log(f_0) + \tau\log(f_1)} \\ &= f_0 \left(\frac{f_1}{f_0} \right)^\tau, \quad \text{for } \tau \in [0, 1]. \end{aligned} \quad (19)$$

This is made precise in the following theorem.

Proposition 7: Consider the Riemannian structured on \mathcal{F} endowed by $g_f(\cdot)$. Given any two spectral density functions $f_0, f_1 \in \mathcal{F}$, any continuous function $\alpha(\tau)$ which satisfies $\alpha(0) = \alpha(1) = 1$, and any continuous non-decreasing function $\beta(\tau)$ which satisfies $\beta(0) = 0$ and $\beta(1) = 1$, then

$$\alpha(\tau)f_0 \left(\frac{f_1}{f_0} \right)^{\beta(\tau)}, \quad \text{for } \tau \in [0, 1] \quad (20)$$

is a geodesic connecting the two given densities. Conversely, all possible geodesics between f_0 and f_1 are given by the above expression for a suitable choice of α and β . Moreover, the geodesic distance is

$$d_g(f_0, f_1) = \|\log(f_0) - \log(f_1)\|_{\text{var}}.$$

Proof: Since the path-length (18) only depends on $\dot{f}_\tau(\theta)/f_\tau(\theta)$, we define

$$x_\tau := \log(f_\tau)$$

and express (18) in terms of $\dot{x}_\tau(\theta) = \dot{f}_\tau(\theta)/f_\tau(\theta)$, i.e.,

$$\ell(f_\tau : 0, 1) := \int_0^1 \|\dot{x}\|_{\text{var}} d\tau. \quad (21)$$

The requirement that the end points of f_τ coincide with f_0 and f_1 readily translates into boundary conditions for x_τ , namely $x_0 = \log f_0$ and $x_1 = \log f_1$. The task of finding extremals of such integrals requires solving Euler–Lagrange equations for the “rescaled” path x_τ . The Lagrangian

$$L(x_\tau, \dot{x}_\tau, \tau) := \|\dot{x}\|_{\text{var}}$$

in (21) only depends on \dot{x}_τ . Therefore, $\partial L/\partial x_\tau = 0$ and the Euler–Lagrange equations

$$\frac{\partial L}{\partial x_\tau} - \frac{\partial}{\partial \tau} \frac{\partial L}{\partial \dot{x}_\tau} = 0$$

simplify to $\partial L/\partial \dot{x}_\tau$ being independent of τ . Since \dot{x}_τ enters in L through an integral over θ , the partial derivative with respect to \dot{x}_τ is infinitesimal. Thus, we write

$$\frac{\partial L}{\partial \dot{x}_\tau} = v(\theta)d\theta$$

which is independent of τ , as we just explained. Since

$$\begin{aligned} \frac{\partial L}{\partial \dot{x}_\tau} &= \frac{1}{2} \frac{1}{\sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} (\dot{x}_\tau(\theta))^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{x}_\tau(\theta) d\theta \right)^2}} \\ &\quad \times \frac{\partial}{\partial \dot{x}_\tau} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (\dot{x}_\tau(\theta))^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{x}_\tau(\theta) d\theta \right)^2 \right) \end{aligned}$$

where the latter term produces again a differential in θ , it follows that

$$\frac{\dot{x}_\tau(\theta) - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{x}_\tau(\theta) d\theta \right)}{\|\dot{x}\|_{\text{var}}} = v(\theta). \quad (22)$$

We first show that

$$\dot{x}_\tau(\theta) = A(\tau) + B(\tau)v(\theta) \quad (23)$$

for suitable functions A, B of τ . Set $A(\tau) = (1/2\pi) \times \int_{-\pi}^{\pi} \dot{x}_\tau(\theta) d\theta$ and $Y(\tau, \theta) = \dot{x}_\tau(\theta) - A(\tau)$. Substituting in (22), we obtain that

$$\frac{Y(\tau, \theta)}{\sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\tau, \theta)^2 d\theta}} = v(\theta).$$

It therefore follows that $\dot{x}_\tau(\theta)$ satisfies (23) for $A(\tau)$ as above and $B(\tau) = \sqrt{(1/2\pi) \int_{-\pi}^{\pi} Y(\tau, \theta)^2 d\theta}$. Hence

$$\begin{aligned} x_\tau(\theta) &= x_0(\theta) + \int_0^\tau \dot{x}_t(\theta) dt \\ &= x_0(\theta) + \int_0^\tau A(t) dt + v(\theta) \int_0^\tau B(t) dt. \end{aligned}$$

Letting $x_0 = \log(f_0)$, $\alpha(\tau) = \exp(\int_0^\tau A(t) dt)$, and $\beta(\tau) = \int_0^\tau B(t) dt$ we have that

$$f_\tau(\theta) = \alpha(\tau)f_0(\theta)v(\theta)^{\beta(\tau)}.$$

Matching the required boundary condition at $\tau = 1$, we have $f_1 = \alpha(1)f_0v^{\beta(1)}$. Without loss of generality, we normalize $\alpha(1) = 1$ and $\beta(1) = 1$, which gives

$$v(\theta) = f_1(\theta)/f_0(\theta).$$

Thus, it is necessary that geodesics are as in (20). Condition (20) is also sufficient since there are no other extremal paths, and they all have identical path length as follows:

$$\begin{aligned}
 \ell(f_\tau : 0, 1) &= \int_0^1 \left\| \frac{\dot{f}_\tau}{f_\tau} \right\|_{\text{var}} d\tau \\
 &= \int_0^1 \left\| \frac{\dot{\alpha}}{\alpha} + \log(v)\dot{\beta} \right\|_{\text{var}} d\tau \\
 &= \int_0^1 \left\| \log(v)\dot{\beta} \right\|_{\text{var}} d\tau \\
 &= \int_0^1 \|\log(v)\|_{\text{var}} |\dot{\beta}| d\tau \\
 &= \|\log(v)\|_{\text{var}}
 \end{aligned}$$

where in step 2 we use the fact that $\dot{\alpha}/\alpha$ does not depend on θ , and in step 3, the assumption that $\dot{\beta} \geq 0$. If β is not monotonically nondecreasing, the path length

$$\int_0^1 \|\log(v)\|_{\text{var}} |\dot{\beta}| d\tau > \int_0^1 \|\log(v)\|_{\text{var}} \dot{\beta} d\tau = \|\log(v)\|_{\text{var}}$$

is not minimal (since it “traces” some parts twice). ■

Remark 8: The geodesic length is a pseudometric on \mathcal{F} . The logarithmic map

$$\log : \mathcal{F} \rightarrow L_2[-\pi, \pi] : f \mapsto \log(f)$$

takes \mathcal{F} into a linear space. It is interesting that the geodesics are mapped, essentially, into straight lines

$$f_0 \left(\frac{f_1}{f_0} \right)^\tau \mapsto \log f_0(\theta) + \tau (\log f_1(\theta) - \log f_0(\theta))$$

modulo scaling of the path parameter τ and an additive constant (with respect to θ) $\log(\alpha(\tau))$. Therefore, it would be interesting to consider other standard metrics/norms on $\{\log(f) : f \in \mathcal{F}\}$ such as the 2-norm, and study their significance from a signal-processing point of view.

V. DEGRADATION OF THE SMOOTHING-ERROR VARIANCE AND RIEMANNIAN STRUCTURE

In a way completely analogous to the previous sections, we consider the degradation in the variance of the smoothing error when a smoothing filter is based on a wrong choice between two alternative power spectra. The degradation of performance quantifies the distance between the two power spectra and endows \mathcal{F} with an alternative Riemannian structure.

We begin with $f_1, f_2 \in \mathcal{F}$, and hence f_1^{-1}, f_2^{-1} are integrable (also bounded). Accordingly, we test the optimal smoothing filter based on f_2 against f_1 . As explained in Section II, the f_2 -optimal smoothing filter gives rise to an error $u_0 - \hat{u}_0|_{\text{past \& future}}$ corresponding, via the Kolmogorov

mapping, to $h_{f_2} f_2(\theta)^{-1}$. Hence, the variance of the smoothing error divided by the f_1 -optimal variance is

$$\begin{aligned}
 \rho_{\text{smooth}}(f_1, f_2) &:= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |h_{f_2} f_2(\theta)^{-1}|^2 f_1(\theta) d\theta \right) \frac{1}{h_{f_1}} \\
 &= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f_2(\theta)^{-2} f_1(\theta) d\theta \right) \frac{h_{f_2}^2}{h_{f_1}} \\
 &= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)^2} d\theta \right) \\
 &\quad \times \frac{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f_1(\theta)^{-1} d\theta \right)}{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f_2(\theta)^{-1} d\theta \right)^2}.
 \end{aligned}$$

This can be readily rewritten as follows:

$$\begin{aligned}
 \rho_{\text{smooth}}(f_1, f_2) &= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right)^2 f_1(\theta)^{-1} d\theta \right) \\
 &\quad \times \frac{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)^{-1} d\theta \right)}{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right) f_1(\theta)^{-1} d\theta \right)^2} \\
 &= \left(\frac{\sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right)^2 d\phi_1(\theta)}}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right) d\phi_1(\theta)} \right)^2 \quad (24)
 \end{aligned}$$

where

$$d\phi_1(\theta) := \frac{f_1(\theta)^{-1} d\theta}{\frac{1}{2\pi} \int_{-\pi}^{\pi} f_1(\theta)^{-1} d\theta}$$

is a normalized measure with variation one. Expression (24) is again of a familiar form. It shows that the degradation is the square of the ratio of the mean square of the fraction f_1/f_2 over its arithmetic mean. These two means, mean square and arithmetic, are weighted by $d\phi_1$, which is, of course, dependent on one of the two arguments. As in the case of $\rho(f_1, f_2)$ in (9), the expression is homogeneous and independent of scaling in either of the two arguments f_1 or f_2 .

Accordingly, we may define as a distance measure

$$\begin{aligned}
 \gamma_{\text{smooth}}(f_1, f_2) &= \rho_{\text{smooth}}(f_1, f_2) - 1 \\
 \delta_{\text{smooth}}(f_1, f_2) &= \log(\rho_{\text{smooth}}(f_1, f_2)).
 \end{aligned}$$

For convenience, we again introduce notation

$$\|\varphi\|_{\text{var}, d\phi}^2 := \frac{1}{2\pi} \int_{-\pi}^{\pi} (\varphi(\theta))^2 d\phi(\theta) - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\theta) d\phi(\theta) \right)^2 \quad (25)$$

and note that, as before, $\|\varphi + c\|_{\text{var}, d\phi} = \|\varphi\|_{\text{var}, d\phi}$ for any c that is independent of θ .

Proposition 9: Let $f, f + \Delta \in \mathcal{F}$, where Δ such that $|\Delta/f| < 1$. Then

$$\gamma_{\text{smooth}}(f + \Delta, f) = \delta_{\text{smooth}}(f + \Delta, f) + O(\|\Delta/f\|^3) \quad (26)$$

$$= \left\| \frac{\Delta}{f} \right\|_{\text{var}, d\phi_1} + O(\|\Delta/f\|^3). \quad (27)$$

Proof: It is completely analogous to the proof of Proposition 5; hence, it is omitted. ■

In view of the above, it is natural to define a corresponding Riemannian metric.

Definition 10: For any $f \in \mathcal{F}$, we define the Riemannian metric

$$g_{f,d\phi} : \mathcal{D} \rightarrow \mathbb{R}_+ : \Delta \mapsto g_f(\Delta) := \left\| \frac{\Delta}{f} \right\|_{\text{var}, d\phi}^2 \quad (28)$$

where $d\phi = (f(\theta)^{-1}/((1/2\pi)\int_{-\pi}^{\pi} f(\theta)^{-1}d\theta))d\theta$.

The characterization of geodesics and geodesic distances appears substantially more involved than in the case of the prediction problem and will be left for future investigations.

VI. CONTRAST WITH INFORMATION GEOMETRY

We believe that it is instructive to contrast the material presented herein in the backdrop of information geometry. Therefore, below, we summarize certain of the key concepts. The monograph of Amari and Nagaoka [2] contains an overview of the development of the subject during the past 50 years.

We focus on the differential geometry of the finite-dimensional probability simplex

$$\mathcal{P} := \left\{ p(k); k \in \{1, \dots, n\} p(k) > 0, \sum_k p(k) = 1 \right\}.$$

The optimal codelength for a random source of independent symbols generated according to $p_1 \in \mathcal{P}$ is $-\sum_k p_1(k) \log(p_1(k))$ [9, Sec. 5.4]. When the code is based on the wrong choice between two alternatives $p_1, p_2 \in \mathcal{P}$, i.e., when the code is designed based on p_2 while the symbols are generated according to p_1 , then the degradation of code length efficiency is given by the difference between actual and ideal code lengths

$$K(p_1, p_2) := \left(-\sum_k p_1(k) \log(p_2(k)) \right) - \left(-\sum_k p_1(k) \log(p_1(k)) \right) \quad (29)$$

which is precisely the Kullback–Leibler divergence. Infinitesimal perturbations from a given $p \in \mathcal{P}$ give rise to a Riemannian metric

$$\begin{aligned} K(p, p + \Delta) &\simeq g_{p, \text{Fisher}}(\Delta) = \sum_k \frac{\Delta(k)^2}{p(k)} \\ &= \sum_k \left(\frac{\Delta(k)}{p(k)} \right)^2 p(k) \\ &=: \left\| \frac{\Delta}{p} \right\|_p^2 \end{aligned} \quad (30)$$

which is known as the Fisher information metric. The map

$$p(k) \mapsto u(k) := 2\sqrt{p(k)} \quad (31)$$

which takes \mathcal{P} onto the first orthant of the sphere $\sum_k u(k)^2 = 4$, turns out to correspond distances measured by the Fisher metric on the probability simplex, to Euclidean distances (e.g., see [1] and [7]). Geodesics on \mathcal{P} correspond to great circles, and geodesic distances in the Fisher metric on the probability simplex correspond to lengths of arcs on the sphere. Then, the Bhattacharyya distance [4] $B(p_1, p_2) := \sum_k \sqrt{p_1(k)p_2(k)}$, which is the cosine of the geodesic arc between the two image points under (31), the geodesic arc, the angle $\arccos(B(p_1, p_2))$, and the Hellinger discrimination [14] $H(p_1, p_2) := \sum_k (\sqrt{p_1(k)} - \sqrt{p_2(k)})^2$, all represent equivalent metrics induced by the geometry of the Fisher information metric.

It is interesting to contrast the metrics introduced earlier

$$g_f(\Delta) = \left\| \frac{\Delta}{f} \right\|_{\text{var}}^2 \quad \text{and} \quad g_{f,d\phi}(\Delta) = \left\| \frac{\Delta}{f} \right\|_{\text{var}, d\phi}^2$$

with the Fisher information metric

$$g_{p, \text{Fisher}}(\Delta) = \left\| \frac{\Delta}{p} \right\|_p^2.$$

Both g_f and $g_{p, \text{Fisher}}$ admit a rather explicit characterization of geodesics and explicitly computable geodesic distances, albeit quite distinct ones. However, with regard to the defining mathematical expressions, there appears to be a closer affinity between $g_{p, \text{Fisher}}$ and $g_{f,d\phi}$; yet, an explicit characterization of geodesics/distances for the latter is still open. The Kullback–Leibler divergence (29), on the other hand, is analogous to the degradation of predictive-error variance (8) of Section III and of smoothing-error variance of Section V. In all cases, these quantities represent degradation of performance of a certain design when the wrong choice between two alternatives is made. Accordingly, they endow the manifold of the respective distributions with a Riemannian metric and appropriate geodesic distances (metrics or pseudometrics, respectively).

Perhaps, a crowning result of information geometry is Chentsov characterization of the Fisher information metric as the unique metric for which stochastic maps are contractive [7], [8]. The possible relationship of $g_f, g_{f,d\phi}$ with classes of natural transformations on spectral density functions, analogous to stochastic maps in the probabilistic context, remains to be studied.

VII. CONCLUDING REMARKS

The manifold \mathcal{F} of power spectral density functions can be endowed with a variety of Riemannian metrics. Naturally, such metrics are helpful in quantifying distances. We presented two alternatives g_f and $g_{f,d\phi}$, which are motivated by a prediction and a smoothing problem, respectively. The geometry induced

by g_f , in particular, is especially simple because, in essence, the logarithmic map

$$f \mapsto \log f$$

takes geodesics to straight lines. This fact raises the possibility that logarithmic m -simplices, e.g.,

$$\mathcal{S}_m := \left\{ f = f_1^{\tau_1} f_2^{\tau_2} \cdots f_m^{\tau_m} : \sum_{i=1}^m \tau_i = 1, \right. \\ \left. \text{and } \tau_i \geq 0, i = 1, \dots, m \right\}$$

may be useful in modeling spectral uncertainty and in modeling families of spectra with a certain affinity. This is reminiscent of geometric mixtures that appear in geometric programming. From a computational viewpoint, logarithmic simplices are certainly attractive since the geodesic distance of a given power spectrum $f_0 \in \mathcal{F}$ from such a simplex \mathcal{S}_m

$$\min_{\tau_1, \dots, \tau_m} \left\{ \left\| \tau_1 \log \left(\frac{f_1}{f_0} \right) + \cdots + \tau_m \log \left(\frac{f_m}{f_0} \right) \right\|_{\text{var}} : \sum_{i=1}^m \tau_i = 1, \right. \\ \left. \text{and } \tau_i \geq 0, i = 1, \dots, m \right\}$$

is readily solvable with standard tools—being an ordinary least-squares problem with inequality constraints. The value of a differential geometric viewpoint and of a natural metric appears especially pertinent in hypothesis testing and, in general, for applications requiring symmetric distances for comparing alternatives and evaluating models of time series.

REFERENCES

[1] S. Amari, *Differential-Geometrical Methods in Statistics, Lecture notes in Statistics*. Berlin, Germany: Springer-Verlag, 1985.
 [2] S. Amari and H. Nagaoka, "Methods of information geometry," *Trans. Math. Monographs*, vol. 191, Amer. Math. Soc., 2000.
 [3] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Stat. Soc.*, vol. 28, pp. 131–142, 1966.
 [4] 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.
 [5] E. F. Beckenbach and R. Bellman, *Inequalities*. Berlin, Germany: Springer-Verlag, 1965, 198 pp.
 [6] 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," *U.S.S.R. Comput. Math. Math. Phys.*, vol. 7, pp. 200–217, 1967.
 [7] L. L. Campbell, "The relation between information theory and the differential geometry approach to statistics," *Inf. Sci.*, vol. 35, pp. 199–210, 1985.

[8] N. N. Chentsov, "Statistical decision rules and optimal inference," in *Translations of Mathematical Monographs*. Providence, RI: AMS, 1982.
 [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
 [10] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Probab.*, vol. 19, no. 4, pp. 2032–2066, 1991.
 [11] T. T. Georgiou, "The maximum entropy principle in the absence of a time arrow: Fractional-pole models," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, Aug. 2007, to be published.
 [12] U. Grenander and G. Szegő, *Toeplitz Forms and their Applications*. New York: Chelsea, 1958.
 [13] S. Haykin, *Nonlinear Methods of Spectral Analysis*. New York: Springer-Verlag, 1979, 247 pp.
 [14] E. Hellinger, "Neue Begründung der theorie der quadratischen formen von unendlichen vielen ver anderlichen," *J. Reine Aug. Math.*, vol. 136, pp. 210–271, 1909.
 [15] R. Gray, A. Buzo, A. Gray, and Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, pp. 367–376, Aug. 1980.
 [16] A. Gray, Jr. and J. Markel, "A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 3, pp. 207–217, Jun. 1974.
 [17] P. Masani, "Recent Trends in Multivariate Prediction Theory," in *Multivariate Analysis*, P. R. Krishnaiah, Ed. New York: Academic, 1966, pp. 351–382.
 [18] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1987.
 [19] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2005.
 [20] G. Szegő, "Über die randwerten einer analytischen functionen," *Math. Ann.*, vol. 84, pp. 232–244, 1921.
 [21] S. R. S. Varadhan, *Probability Theory*. Providence, RI: AMS, 2000.



Tryphon T. Georgiou (F'00) was born in Athens, Greece, on October 18, 1956. He received the Diploma degree in mechanical and electrical engineering from the National Technical University of Athens, Greece, in 1979 and the Ph.D. degree from the University of Florida, Gainesville, in 1983. He served on the faculty of Florida Atlantic University, Boca Raton, from 1983 to 1986 and Iowa State, Ames, from 1986 to 1989. Since 1989, he has been with the University of Minnesota, where he holds the Vincentine Hermes-Luh Chair of Electrical

Engineering. Dr. Georgiou has received the George S. Axelby Outstanding Paper Award of the IEEE Control Systems Society three times, for 1992, 1999, and 2003. In 1992 and in 1999, he received the award for joint work with Prof. M. C. Smith (Cambridge University, Cambridge, U.K.) and, in 2003, for joint work with Prof. C. Byrnes (Washington University, St. Louis, MO) and Prof. A. Lindquist (KTH, Stockholm, Sweden). He has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the *SIAM Journal on Control and Optimization*, the *Systems and Control Letters*, and he has served on the Board of Governors of the Control Systems Society of the IEEE.