UC Irvine
UC Irvine Previously Published Works

Title
Distances between power spectral densities

Permalink
https://escholarship.org/uc/item/9hn1w3sm

Author
Georgiou, Tryphon T

Publication Date
2006-07-01

Copyright Information

Peer reviewed

Distances between power spectral densities[*]

Tryphon T. Georgiou[†]

Abstract

We present several natural notions of distance between spectral density functions of (discrete-time) random processes. They are motivated by certain filtering problems. First we quantify the degradation of performance of a predictor which is designed for a particular spectral density function and then it is used to predict the values of a random process having a different spectral density. The logarithm of the ratio between the variance of the error, over the corresponding minimal (optimal) variance, produces a measure of distance between the two power spectra with several desirable properties. Analogous quantities based on smoothing problems produce alternative distances and suggest a class of measures based on fractions of generalized means of ratios of power spectral densities. These distance measures endow the manifold of spectral density functions with a (pseudo) Riemannian metric. We pursue one of the possible options for a distance measure, characterize the relevant geodesics, and compute corresponding distances.

Index Terms

Power spectral density functions, distance measures.

I. INTRODUCTION

DUE 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 which attempt to acknowledge the structure of power spectral density functions as a positive cone. The most most well known are the Kullback-Leibler divergence which originates in hypothesis testing and in Bayes' estimation, the Itakura-Saito distance which originates in speech analysis —both belonging to Bregman class ([13], [4], [12]), the Bhattacharyya distance [2], and the Ali-Silvey class of divergences [1]. 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 certain properties (as in the case of Bregman and Ali-Silvey classes). The purpose of this work is to introduce certain new notions of distance which are rooted in filtering theory and provide intrinsic distance measures between any two *power spectral density* functions.

Our starting point is a prediction problem. We select an optimal predictive filter for an underlying random process based on the assumption that the process has a given power spectral density $f_1(\theta)$. We then evaluate the performance of such a filter against a second power spectral density $f_2(\theta)$ —which may be thought of as the spectral density function of the "actual" random process. The relative degradation of performance (i.e., variance of the prediction error) quantifies a mismatch between the two functions. Interestingly, it turns out to be equal to the ratio of the arithmetic over the geometric mean of the fraction of the two power spectra. The logarithm of the relative degradation serves a distance measure.

Infinitesimal analysis suggests a pseudo-Riemannian metric on the manifold of power spectral density functions. The presence of such a metric suggests that geodesic distances may be used to quantify divergence between power spectra. Indeed, a characterization of geodesics is provided, and certain logarithmic intervals are shown to satisfy the condition. The length of such intervals connecting two power spectral densities provides yet another notion of distance between the two.

[†]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455; `tryphon@ece.umn.edu`.

An identical approach based on the degradation of performance of smoothing filters leads to other expressions which, equally well, quantify divergence between power spectral densities. Two observations appear to be universal. First that the mismatch between the "shapes" of spectral density functions is what turns out to be important. This is quantified by how far the ratio f_1/f_2 of the two spectral densities is from being constant across frequencies. The ratio of spectral density functions is reminiscent of the likelihood ratio in probability theory.

The second observation is that all of the distance measures that we encountered, in essence, they compare different means (i.e., arithmetic, geometric, and harmonic, possibly, weighted) of the ratio of the two spectral density functions or of their logarithms. It is quite standard, that e.g., the argithmetic and the geometric means coincide only when the ratio is constant and have a gap otherwise. The same applies to a wider family of generalized means. Thus, this observation suggests a much larger class of possible alternatives: quantify the divergence between (the "shape" of) two density functions using the gap between two generalized means of their ratio, or by the slackness of Jensen-type of inequalities involving this ratio. The underlying mathematical construct appears quite distinct from those utilized in defining the Bregman and the Ali-Silvey classes of distance measures. Furthermore, the mathematical construct is deeply rooted in prediction theory and, at least in certain cases, can be motivated as quantifying degradation of performance as we explained earlier.

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 $R_0, R_1, R_2, \ldots$ its sequence of autocorrelation samples and by $d\mu(\theta)$ its power spectrum. Thus, $R_k := \mathcal{E}\{u_\ell u^*_{\ell+k}\} = R^*_{-k}$ and

$$R_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jk\theta} d\mu(\theta), \ \text{for } k \in \mathbb{Z},$$

while $\mathcal{E}$ denotes expectation and "*" denotes complex conjugation. We are interested in quadratic optimization problems with respect to the usual inner product

$$\langle \sum_k a_k u_k, \sum_\ell b_\ell u_\ell \rangle \ := \ \mathcal{E}\{(\sum_k a_k u_k)(\sum_\ell b_\ell u_\ell)^*\}$$
$$= \ \sum_{k,\ell} a_k R_{k-\ell} b^*_\ell. \tag{1}$$

The closure of $\mathrm{span}\{u_k \ : \ k \in \mathbb{Z}\}$, 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 $d\mu(\theta)$ 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 $b(\theta) = \sum_\ell b_\ell e^{j\ell\theta}$. Further, the correspondence

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

is a Hilbert space isomorphism (see [14]). 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

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

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

$$\hat{u}_{0|\text{past \& future}} := \sum_{k \neq 0} \beta_k u_{-k} \tag{3}$$

is simply

$$\mathcal{E}\{|u_0 - \hat{u}_{0|\text{past \& future}}|^2\} = \|1 - \sum_{k \neq 0} \beta_k e^{jk\theta}\|_{d\mu}^2. \tag{4}$$

In general, the power spectrum $d\mu$ is a bounded nonnegative measure on $[-\pi, \pi)$ and admits a decomposition $d\mu = d\mu_s + f d\theta$ with $d\mu_s$ a singular measure and $f d\theta$ 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 Szegö-Kolmogorov formula stated below (see [17] and also, [10, page 183], [18, Chapter 6], [11], [16]).

Theorem 1: With $d\mu = d\mu_s + f d\theta$ as above

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

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 non-deterministic in the sense of Kolmogorov. In this case, it can be shown that

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

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

$$a_f(z) = 1 + a_1 z + a_2 z^2 + \ldots$$

is analytic in the unit disc $\mathbb{D} := \{z : |z| < 1\}$ and its radial limits are square integrable (see [15]). Then, the linear combination

$$\hat{u}_{0|\text{past}} := \sum_{k>0} (-a_k) u_{-k} \tag{5}$$

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

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

Analogous expressions exist for the optimal smoothing error and the corresponding smoothing filter which uses both past and future values of u_ℓ. 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.

Theorem 2: (see [9]) With $d\mu = d\mu_s + f d\theta$ as above

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

$$=: h_f \tag{6}$$

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 nondeterministic in the sense that past and future specify the present which can be estimated with zero variance. In this case (see [9])

$$\begin{aligned}
b_f(\theta) &= \ldots + b_2 e^{-2j\theta} + b_{-1} e^{-j\theta} + 1 + \\
&\quad + b_1 e^{j\theta} + b_2 e^{2j\theta} + \ldots \\
&= h_f f(\theta)^{-1}
\end{aligned}$$

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

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

III. Degradation of the Prediction Error Variance

We now consider two distinct spectral density functions f_1, 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 serves as a measure of mismatch between the two spectral densities (the one which 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 convex on logarithmic intervals and has a number of distance-like properties, short of being a metric.

Let us assume that both $\log f_1, \log f_2 \in L_1[-\pi, \pi)$ and hence, that

$$f_i(\theta) = \frac{g_{f_i}}{|a_{f_i}(e^{j\theta})|^2}$$

for corresponding outer H_2-functions $a_{f_i}(z)$ normalized as before so that $a_{f_i}(0) = 1$, for $i \in \{1, 2\}$. Obviously,

$$g_{f_i} = \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f_i(\theta)) d\theta\right)$$

denotes the geometric mean of f_i as before, for $i \in \{1, 2\}$. These expressions represent the least variances when the predictor is chosen on the basis of the correct spectral density function. If however, 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 turns out to be

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

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

$$
\begin{aligned}
\rho_{a/g}(f_1, f_2) \quad &:= \quad \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta \right) \frac{g_{f_2}}{g_{f_1}} \tag{7} \\
&= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta \right) \frac{\exp(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_2(\theta) d\theta)}{\exp(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_1(\theta) d\theta)} \\
&= \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta \right) \frac{1}{\exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\frac{f_1(\theta)}{f_2(\theta)}) d\theta \right)}. \tag{8}
\end{aligned}
$$

This is the ratio of the *arithmetic mean* over the *geometric mean* of the fraction f_1/f_2 of the two spectral density functions. The expression $\rho_{a/g}(f_1, f_2)$ is not symmetric in the two arguments. The subscript "$_{a/g}$" signifies ratio of arithmetic over geometric means.

The logarithm of $\rho_{a/g}(f_1, f_2)$ is nonnegative and defines a notion of distance between rays of density functions. Henceforth, we denote this logarithm by

$$
\begin{aligned}
\delta_{a/g}(f_1, f_2) \quad &:= \quad \log\left(\rho_{a/g}(f_1, f_2) \right) \tag{9} \\
&= \quad \log\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_2(\theta)} d\theta \right) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(\frac{f_1(\theta)}{f_2(\theta)} \right) d\theta.
\end{aligned}
$$

Alternatively, we can view the above as slackness of a Jensen-type inequality.

Before we discuss key properties of $\delta_{a/g}$, we introduce a natural class of paths connecting density functions: for any two density functions f_a, f_b,

$$
f_{\tau, a, b} := f_a^{1-\tau} f_b^\tau, \text{ for } \tau \in [0, 1],
$$

defines a *logarithmic interval* between f_a and f_b. The terminology stems from the fact that whenever the needed logarithms exist,

$$
f_{\tau, a, b} = e^{(1-\tau)\log(f_a) + \tau \log(f_b)}, \text{ for } \tau \in [0, 1].
$$

Later on we will see that these represent geodesics on the manifold of density functions with respect to an induced pseudo-Riemannian metric.

Proposition 3: Let f_i, $i \in \{1, 2, 3\}$ represent density functions defined on $[-\pi, \pi)$. The following hold:
  (i) $\delta_{a/g}(f_1, f_2) \in \mathbb{R}_+ \cup \{\infty\}$.
  (ii) $\delta_{a/g}(f_1, f_2) = 0 \Leftrightarrow f_1(\theta)/f_2(\theta)$ is constant.
  (iii) $\delta_{a/g}(f_1, f_{\tau, 1, b})$ is monotonically increasing
        for $\tau \in [0, 1]$ and $b \in \{2, 3\}$.
  (iv) $\delta_{a/g}(f_1, f_{\tau, 2, 3})$ is convex in τ.

Proof: Claims (i-ii) follow from the fact that the arithmetic mean of a function always exceeds the geometric mean, and that they are equal whenever a function is constant. In particular, the ordering, as to which is larger, follows from Jensen's inequality

$$
\log\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) d\theta \right) \geq \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(f(\theta) \right) d\theta
$$

for any $f(\theta) \geq 0$. The fact that the two are equal only when $f(\theta)$ is constant can be obtained easily using a variational argument. Then (i-ii) follow, when we substitute $f = f_1/f_2$ and then take the logarithm.

Next we show (iv) and use it to deduce (iii). Since

$$
\begin{aligned}
\delta(f_1, f_a^{(1-\tau)} f_b^\tau) &= \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a^{(1-\tau)}(\theta) f_b^\tau(\theta)} d\theta \\
&\quad - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(\frac{f_1(\theta)}{f_a^{(1-\tau)}(\theta) f_b^\tau(\theta)} \right) d\theta \\
&= \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_a(\theta)} \right) \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \\
&\quad - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(\frac{f_1(\theta)}{f_a(\theta)} \right) d\theta \\
&\quad - \frac{1}{2\pi} \int_{-\pi}^{\pi} \tau \log \left(\frac{f_a(\theta)}{f_b(\theta)} \right) d\theta,
\end{aligned}
$$

the derivative with respect to τ becomes

$$
\begin{aligned}
\frac{d}{d\tau} \delta(f_1, f_a^{(1-\tau)} f_b^\tau) &= \frac{1}{\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta} \\
&\quad \times \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\log \frac{f_a(\theta)}{f_b(\theta)} \right) \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \\
&\quad - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(\frac{f_a(\theta)}{f_b(\theta)} \right) d\theta
\end{aligned}
$$

and the second derivative,

$$
\begin{aligned}
\frac{d^2}{d\tau^2} \delta(f_1, f_a^{(1-\tau)} f_b^\tau) &= \frac{-1}{\left(\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \right)^2} \\
&\quad \times \left(\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\log \frac{f_a(\theta)}{f_b(\theta)} \right) \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \right)^2 \\
&\quad + \frac{1}{\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta} \int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\log \frac{f_a(\theta)}{f_b(\theta)} \right)^2 \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta
\end{aligned}
$$

But from Cauchy's inequality we have that

$$
\begin{aligned}
&\left(\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \right) \\
&\quad \times \left(\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\log \frac{f_a(\theta)}{f_b(\theta)} \right)^2 \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \right) \\
&\geq \left(\int_{-\pi}^{\pi} \frac{f_1(\theta)}{f_a(\theta)} \left(\log \frac{f_a(\theta)}{f_b(\theta)} \right) \left(\frac{f_a(\theta)}{f_b(\theta)} \right)^\tau d\theta \right)^2.
\end{aligned}
$$

Hence, the second derivative is nonnegative. Claim (iv) is seen to hold true after we set $a = 2$ and $b = 3$. To establish claim (iii) set $a = 1$ and $b = 2$ in the above. Then,

$$
\frac{d^2}{d\tau^2} \delta(f_1, f_1^{(1-\tau)} f_2^\tau) \geq 0.
$$

But $\delta(f_1, f_1^{(1-\tau)} f_2^\tau) = 0$ for $\tau = 0$, and $\delta(f_1, f_1^{(1-\tau)} f_2^\tau) \geq 0$ for all τ. Hence, the derivative at $\tau = 0$ must be nonnegative and $\delta(f_1, f_1^{(1-\tau)} f_1^\tau)$ must increase as $\tau \nearrow 1$. This completes the proof of (iii), and the proof of the proposition. ■

Since $\delta_{a/g}(f_1, f_2)$ is not symmetric in its arguments, it is quite natural to consider the symmetrized version

$$\delta(f_1, f_2) := \delta_{a/g}(f_1, f_2) + \delta_{a/g}(f_2, f_1).$$

All properties listed in Proposition 3 hold true for $\delta(f_1, f_2)$ as well. Furthermore, interestingly,

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

which is the *logarithm of the ratio of the arithmetic mean over the harmonic mean* of the "likelihood" fraction f_1/f_2. Again, the distance of this ratio from one quantifies how far f_1/f_2 is from being constant. We now summarizing the claimed properties of $\delta(\cdot, \cdot)$.

Proposition 4: Let f_i, $i \in \{1, 2, 3\}$, density functions on $[-\pi, \pi)$. The following hold:
 (i) $\delta(f_1, f_2) \in \mathbb{R}_+ \cup \{\infty\}$.
 (ii) $\delta(f_1, f_2) = 0 \Leftrightarrow f_1(\theta)/f_2(\theta)$ is constant.
 (iii) $\delta(f_1, f_{\tau,1,b})$ is monotonically increasing
       for $\tau \in [0, 1]$ and $b \in \{2, 3\}$.
 (iv) $\delta(f_1, f_{\tau,2,3})$ is convex in τ.

Proof: Properties (i), (ii), and (iv) are a direct consequence of the corresponding properties given in Proposition 3 for $\delta_{a/g}(\cdot, \cdot)$. Property (iii) on the other hand follows as before from (iv) and the fact that the derivative of $\delta(f_1, f_{\tau,1,b})$ at $\tau = 0$ is zero. ■

IV. AN EXAMPLE

In order to illustrate the quantitative behavior of these measures, we consider three specific power spectra labeled f_1, f_2, f_3, as before. These are shown in Figure 1. We then consider the triangle formed with those power spectra as vertices and connected using logarithmic intervals. The interior of the triangle is similarly sampled at logarithmically placed points. In essence, we consider the family of power spectral densities

$$f_1^{(1-\tau)}\left(f_2^{(1-\sigma)}f_3^{\sigma}\right)^{\tau} \quad \text{for } \tau, \sigma \in [0, 1]. \tag{10}$$

For each value of τ, σ (sampled appropriately), we evaluate $\delta_{a/g}(f_1, f)$, $\delta(f_1, f)$ and compare these to the Kullback-Leibler divergence between the suitably normalized functions

$$\hat{f}_k(\theta) = \frac{f_k(\theta)}{\frac{1}{2\pi}\int_{-\pi}^{\pi}f_k d\theta}, \quad k \in \{1, 2, 3\}.$$

The normalization is necessary if the Kullback-Leibler divergence is to have properties of a distance measure (i.e., nonnegative when its arguments are different, etc.). Thus, we denote

$$
\begin{aligned}
\delta_{\mathrm{KL}}(f_1, f_2) &:= \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{f}_1(\theta)\log\left(\frac{\hat{f}_1(\theta)}{\hat{f}_2(\theta)}\right)d\theta \\
&= \frac{1}{\frac{1}{2\pi}\int_{-\pi}^{\pi}f_1(\theta)d\theta}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}f_1(\theta)\left(\log(\frac{f_1(\theta)}{f_2(\theta)}) - \log\frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}f_1(\theta)d\theta}{\frac{1}{2\pi}\int_{-\pi}^{\pi}f_2(\theta)d\theta}\right)d\theta\right).
\end{aligned} \tag{11}
$$

The set of power spectra in (10) is thought of as a set of points forming an equilateral triangle, conceptually sitting on the xy-plane. Then, the vertical axis represents distance from f_1, measured using these three

alternative measures. The corresponding surfaces are drawn in Figures 2-4. The three power spectra used are as follows:

$$
\begin{aligned}
f_1(\theta) &= \left| \frac{(z - .99)}{(z^2 + .6z + .99)} \right|^2_{z=e^{j\theta}} \\
f_2(\theta) &= \left| \frac{1}{(z^2 - .3z + .99)} \right|^2_{z=e^{j\theta}} \\
f_3(\theta) &= \left| \frac{(z + .9)(z^2 + .6z + .99)}{(z^2 + .9z + .99)(z^2 + .9z + .99)} \right|^2_{z=e^{j\theta}}
\end{aligned}
$$

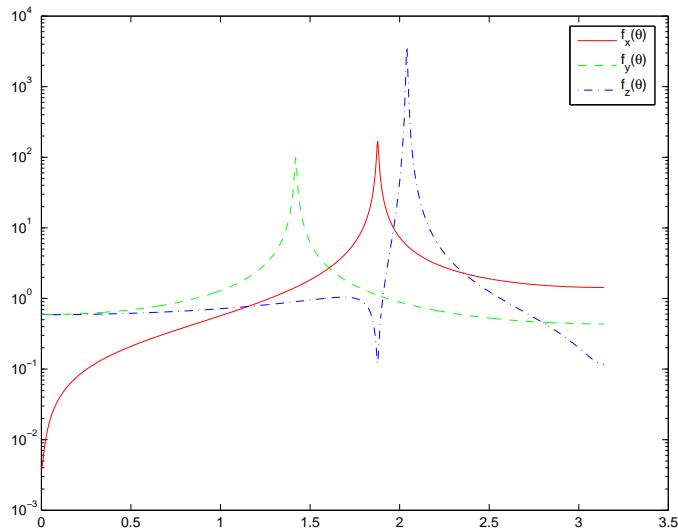

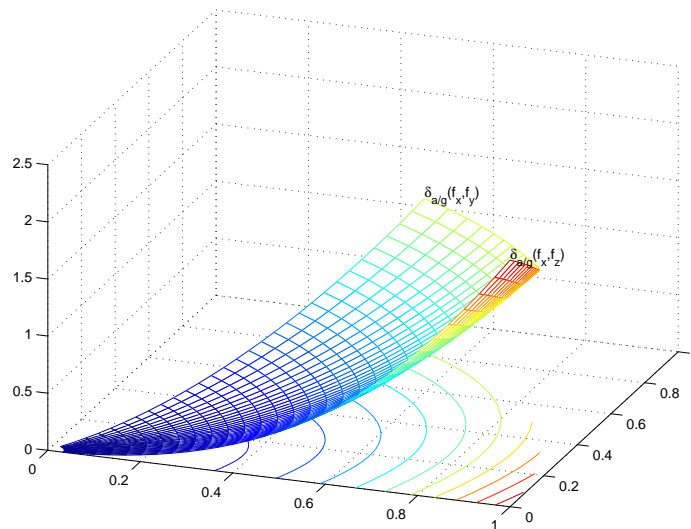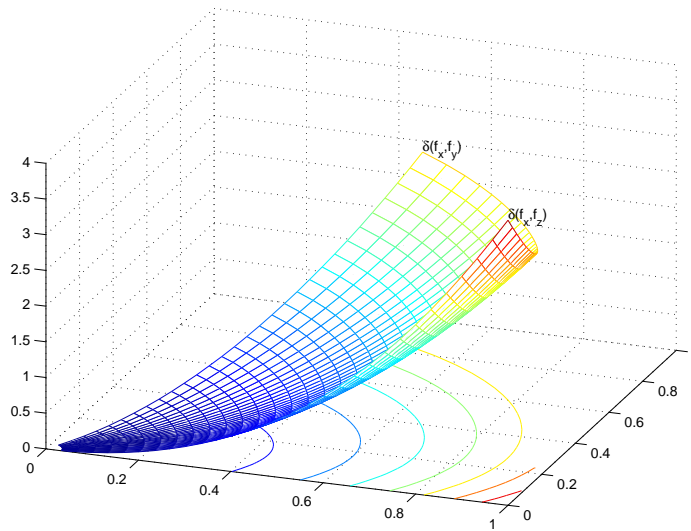Fig. 1.   Three power spectral densities f_1 (–), f_2 (- -), f_3 (-··-)
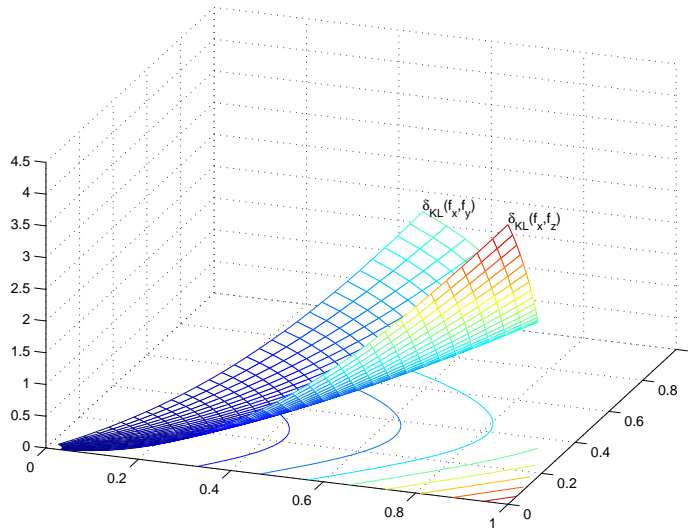

Fig. 2. $\delta_{a/g}(f_1, f_1^{1-\tau} f_2^{\tau - \sigma\tau} f_3^{\sigma\tau})$

There appears to be little qualitative difference between $\delta(\cdot, \cdot)$, $\delta_{a/g}(\cdot, \cdot)$, and $\delta_{\mathrm{KL}}(\cdot, \cdot)$. They are also quite similar in that it is easy to calculate functional forms for minimizers of either of these distance measures

Fig. 3. $\delta(f_1, f_1^{1-\tau} f_2^{\tau - \sigma\tau} f_3^{\sigma\tau})$

Fig. 4. $\delta_{\mathrm{KL}}(f_1, f_1^{1-\tau} f_2^{\tau - \sigma\tau} f_3^{\sigma\tau})$

under moment constraints (see Section V below). Hence, it is important to undercore that $\delta_{\mathrm{KL}}(\cdot, \cdot)$ lacks an intrinsic interpretation as a distance measure between power spectra, in contrast to $\delta(\cdot, \cdot)$ which therefore may be preferable for exactly that reason.

V. FUNCTIONAL FORM OF MINIMIZERS IN MOMENT PROBLEMS

A large class of spectral analysis problems is typified by the trigonometric moment problem where a power spectral density is sought to match a partial sequence of autocorrelation samples, i.e., a positive function f is sought such that

$$R_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jk\theta} f(\theta) d\theta, \text{ for } k \in \{0, 1, \dots, n\}, \tag{12}$$

see e.g., [10], [5], [11], [6], [12]. Since, in general, the family of consistent f's is large, a particular one is chosen "closest" to a given "prior". Maximum entropy spectral analysis, for instance, can be interpreted as seeking the spectral density closest in the Kullback-Leibler sense to one which is flat, i.e., the prior in

this case is the power spectral density of white noise (see e.g., [12]). In the same spirit we may pose the problem of seeking f closest to f_{prior} in the sense of minimizing e.g., $\delta_{a/g}(f, f_{\text{prior}})$ and subject to the moment constraints (12).

To this end, as usual, we introduce Lagrange multipliers λ_k ($k \in \{0, 1, \ldots, n\}$) and form the Lagrangian

$$\mathcal{L}(f, \lambda_0, \ldots, \lambda_n) := \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\theta)}{f_{\text{prior}}(\theta)} d\theta - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \frac{f(\theta)}{f_{\text{prior}}(\theta)} d\theta + \sum_{k=-n}^{n} \lambda_k \left(R_k - \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jk\theta} f(\theta) d\theta \right).$$

Setting the variation of $\mathcal{L}$ identically to zero for all perturbations of f gives conditions that help identify the functional form of minimizing f's. Briefly,

$$
\begin{aligned}
\delta\mathcal{L} &= \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f + \Delta}{f_{\text{prior}}} d\theta - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \frac{f + \Delta}{f_{\text{prior}}} d\theta \\
&\quad - \log \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f}{f_{\text{prior}}} d\theta + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \frac{f}{f_{\text{prior}}} d\theta - \sum_{k=-n}^{n} \lambda_k \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jk\theta} \Delta(\theta) d\theta \right) \\
&\simeq \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta}{f_{\text{prior}}} d\theta}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f}{f_{\text{prior}}} d\theta} - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta}{f} d\theta - \sum_{k=-n}^{n} \lambda_k \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jk\theta} \Delta d\theta \right)
\end{aligned}
$$

after we eliminate higher order terms. Stationarity conditions require that the above is identically zero for all (small) functions Δ. This leads to

$$\frac{1}{f_{\text{prior}}(\theta) \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)/f_{\text{prior}}(\theta) d\theta} - \frac{1}{f(\theta)} - \sum_{k=-n}^{n} \lambda_k e^{jk\theta} = 0$$

from which we deduce that a minimizing f must be of the form

$$f(\theta) = \frac{\kappa f_{\text{prior}}(\theta)}{1 - \kappa f_{\text{prior}}(\theta) \sum_{k=-n}^{n} \lambda_k e^{jk\theta}} \tag{13}$$

with

$$\kappa = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)/f_{\text{prior}}(\theta) d\theta. \tag{14}$$

Then, values for κ as well as for the Lagrange multipliers must be determined so that f in (13) satisfies (12) and (14) —this can be done for instance using homotopy methods in [7], [8]. It is interesting that when $f_{\text{prior}} = 1$, the minimizer is the same as in the one obtained by applying the maximum entropy principle (e.g., see [12], [11]), i.e., it turns out to be an all-pole spectral density function which is of course uniquely identified by the moment constraints. Evidently, in general, minimizing $\delta_{a/g}(f, f_{\text{prior}})$ gives a different answer than the one obtained by minimizing $\delta_{\text{KL}}(f, f_{\text{prior}})$ or, by minimizing other distances. Yet, all such problems are similar and can be dealt with in two steps. First identify the functional form of a minimizer and then determine values for the coefficients so as to satisfy (12). The latter step requires solving a nonlinear problem in general, and can be approached in a variety of ways (e.g., as in [7], [8]).

VI. RIEMANNIAN METRICS AND GEODESICS

Infinitesimal perturbations about a given power spectral density function, when measured by any of $\delta_{a/g}(\cdot, \cdot)$, $\delta(\cdot, \cdot)$, or $\delta_{\text{KL}}(\cdot, \cdot)$, give rise to nonnegative definite quadratic forms. These forms are in fact nonsingular on directions other than rays emanating from the origin. This is due to the fact that the aforementioned distances do not separate points on such rays while they give nonzero distance otherwise. They thus induce Riemannian metrics on suitably defined manifold of spectral rays. In this section (and the current paper) we focus on the particular metric induced by $\delta(\cdot, \cdot)$, we show how to characterize geodesics, and verify that logarithmic intervals are in fact geodesics.

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 section we define

$$
\begin{aligned}
\mathcal{F} \quad &:= \quad \{f \ : \ f(\theta) \text{ differentiable on } [-\pi, \pi], \\
&\text{with } f(\theta) > 0, \text{ and both } f(\theta), \frac{df(\theta)}{d\theta} \text{ square integrable}\}.
\end{aligned}
$$

With a suitable norm on f and its derivative, $\mathcal{F}$ becomes a (Banach) manifold. We also recall the definition $\|f\|_k := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)| d\theta \right)^{1/k}$ of the k-th norm, applicable to any f on $[0, 2\pi]$ provided the integral exist. Whenever, f is a density function, the absolute value is obsiously unnecessary.

Proposition 5: Let $f, f + \Delta \in \mathcal{F}$ where Δ is a perturbation such that $|\Delta/f| < 1$. Then

$$
\delta(f, f + \Delta) \quad = \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)} \right)^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta \right)^2 + O(\|\Delta/f\|_1^3)
$$

$$
\delta_{\mathrm{a/g}}(f, f + \Delta) \quad = \quad \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)} \right)^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta \right)^2 \right) + O(\|\Delta/f\|_1^3)
$$

$$
\delta_{\mathrm{KL}}(f, f + \Delta) \quad = \quad \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(\Delta(\theta))^2}{f(\theta)} d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \Delta(\theta) d\theta \right)^2 \right) + O(\|\Delta/f\|_1^3).
$$

Here, $O(\|\Delta/f\|_1^3)$ indicates terms of order 3 or higher.

Proof: We prove only the first claim. The other two can be shown in an identical manner. We expand $\delta(f, f + \Delta)$ using the series $\log(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots$, as follows

$$
\begin{aligned}
\delta(f, f + \Delta) \quad &= \quad \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\theta) + \Delta(\theta)}{f(\theta)} d\theta \right) + \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\theta)}{f(\theta) + \Delta(\theta)} d\theta \right) \\
&= \quad \log \left(1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta \right) + \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{1 + \frac{\Delta(\theta)}{f(\theta)}} d\theta \right) \\
&= \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta - \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta \right)^2 + O(\|\frac{\Delta(\theta)}{f(\theta)}\|_1^3) \\
&\quad + \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - \frac{\Delta(\theta)}{f(\theta)} + \left(\frac{\Delta(\theta)}{f(\theta)} \right)^2 + \ldots) d\theta \right) \\
&= \quad \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta - \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta \right)^2 \\
&\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} (-\frac{\Delta(\theta)}{f(\theta)} + \left(\frac{\Delta(\theta)}{f(\theta)} \right)^2) d\theta - \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (-\frac{\Delta(\theta)}{f(\theta)}) d\theta \right)^2 + O(\|\frac{\Delta}{f}\|_1^3)
\end{aligned}
$$

which proves the first claim after canceling and collecting terms. The other two expressions can be shown similarly. Note that $\|\Delta/f\|_k \leq \|\Delta/f\|_1$ for all k, since $|\Delta/f| < 1$. ∎

Since $\delta_{a/g}(\cdot, \cdot)$, $\delta(\cdot, \cdot)$, and $\delta_{\mathrm{KL}}(\cdot, \cdot)$ do not separate power spectra which are scalar multiple of one another, we may consider equivalence classes

$$
(f)_\sim \quad := \quad \{f_1 \in \mathcal{F} \ : \ f_1 = cf, c \in \mathbb{R}_+\}.
$$

obtained from any spectral density function f by a scaling factor, constant across $[-\pi, \pi]$. These can be thought of as rays. They can be identified by pointing to one particular representative. Thus, in particular, the set of rays can be identified with

$$
\mathfrak{F} := \{f \ : \ f \in \mathcal{F}, \ \|f\|_1 = 1\}.
$$

This set can be given the structure of a manifold and thought of as a set of probability density functions on $[-\pi, \pi]$. Alternatively, we can consider spectral densities as belonging to $\mathcal{F}$ and accept the fact that we have a pseudo-Riemannian metric which vanishes along certain directions. This follows from Proposition 5, since $\delta(\cdot, \cdot)$ (and similarly, $\delta_{a/g}$ and δ_{KL}) defines a nonnegative definite quadratic form[1] at each "point" $f \in \mathcal{F}$ via

$$\Delta \mapsto \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)} \right)^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} d\theta \right)^2. \tag{15}$$

A rather standard way to measure distances on manifold is to trace geodesics connecting points, and compute the length of such paths. Thus, it may be of interest to characterize geodesics in our case as well. We refrain from excessively technical jargon and, following the earlier suggestion, simply assume that all integrals exist.

Consider a path f_τ, $\tau \in [0, 1]$, of spectral density functions connecting two given ones, namely f_0 and f_1. Note that f_τ is a function of two arguments, the path parameter τ and the frequency θ —hence, we often write $f_\tau(\theta)$. The length traversed as τ varies from 0 to 1 is simply

$$\ell(f_\tau : 0, 1) \;\; := \;\; \int_0^1 \sqrt{\delta(f_\tau, f_{\tau + d\tau})} \tag{16}$$

$$= \int_0^1 \sqrt{ \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_\tau(\theta) + \dot{f}_\tau(\theta) d\tau}{f_\tau(\theta)} d\theta \right) + \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_\tau(\theta)}{f_\tau(\theta) + \dot{f}_\tau(\theta) d\tau} d\theta \right) }$$

$$= \int_0^1 \sqrt{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\dot{f}_\tau(\theta)}{f_\tau(\theta)} \right)^2 d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\dot{f}_\tau(\theta)}{f_\tau(\theta)} d\theta \right)^2 } \, d\tau. \tag{17}$$

In the last step we eliminated higher order terms in $d\tau$ inside the integral, since those integrate to zero. Here and throughout " · " (dot), as in $\dot{f}$ is used to denote derivative with respect to τ, i.e.,

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

Interestingly, the expression in (17) only depends on $\dot{f}_\tau(\theta)/f_\tau(\theta)$. Thus, if we define

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

then $\dot{x}_\tau(\theta) = \dot{f}_\tau(\theta)/f_\tau(\theta)$ and

$$\ell(f_\tau : 0, 1) \;\; := \;\; \int_0^1 \sqrt{ \frac{1}{2\pi} \int_{-\pi}^{\pi} (\dot{x}_\tau(\theta))^2 \, d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{x}_\tau(\theta) d\theta \right)^2 } \, d\tau. \tag{18}$$

The requirement that the end point 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 leads to Euler-Lagrange equations for the path x_τ. More specifically, the Lagrangian corresponding to (18) is

$$L(x_\tau, \dot{x}_\tau, \tau) := \sqrt{ \frac{1}{2\pi} \int_{-\pi}^{\pi} (\dot{x}_\tau(\theta))^2 \, d\theta - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{x}_\tau(\theta) d\theta \right)^2 }$$

and only depends on $\dot{x}_\tau$. Therefore $\partial L/\partial x_\tau = 0$ and the Euler-Lagrange equations

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

---

[1] If we mod-out scaling and stay on a manifold of "spectral rays", this quadratic form becomes positive definite, and defines a Riemannian metric.

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

$$\frac{\partial L}{\partial \dot{x}_\tau} = \frac{1}{2}\frac{1}{\sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}(\dot{x}_\tau(\theta))^2\,d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\dot{x}_\tau(\theta)d\theta\right)^2}}\frac{\partial}{\partial \dot{x}_\tau}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}(\dot{x}_\tau(\theta))^2\,d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\dot{x}_\tau(\theta)d\theta\right)^2\right)$$

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

$$2\sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}(\dot{x}_\tau(\theta))^2\,d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\dot{x}_\tau(\theta)d\theta\right)^2}\,v(\theta)d\theta = 2\dot{x}_\tau(\theta) - 2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\dot{x}_\tau(\theta)d\theta\right)d\theta.$$

Alternatively,

$$\frac{\dot{x}_\tau(\theta) - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\dot{x}_\tau(\theta)d\theta\right)}{\sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}(\dot{x}_\tau(\theta))^2\,d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\dot{x}_\tau(\theta)d\theta\right)^2}} = v(\theta), \tag{19}$$

which simply says that the variation of $\dot{x}_\tau$ about the mean, as a function of θ, normalized by a "standard deviation"-like quantity must be independent of τ. We summarize our conclusion as follows.

Proposition 6: Given two spectral density functions f_0, f_1, extremal (geodesic) paths f_τ ($\tau \in [0,1]$) connecting the two, in the sense of achieving a local extremal of the path integral $\ell(f_\tau : 0, 1)$, must satisfy (19) for $x_\tau = \log f_\tau$, i.e., the left hand side of (19) must be independent of τ.

Proof: The proof has been established in the arguments leading to the proposition. ∎

We finally verify that logarithmic intervals satisfy (19). This is rather straightforward since, for

$$f_\tau(\theta) = f_0(\theta)\left(\frac{f_1(\theta)}{f_0(\theta)}\right)^\tau,$$

the logarithm

$$\begin{aligned}
x_\tau(\theta) &= \log f_\tau(\theta) \\
&= \log f_0(\theta) + \tau(\log f_1(\theta) - \log f_0(\theta))
\end{aligned}$$

is a linear function of τ and the derivative

$$\begin{aligned}
\dot{x}_\tau(\theta) &= \frac{\partial}{\partial \tau}\log(f_\tau(\theta)) \\
&= \log(f_1(\theta)) - \log(f_0(\theta)) \\
&= x_1(\theta) - x_2(\theta)
\end{aligned}$$

is already independent of τ. The ratio f_1/f_0 plays a rôle analogous to the likelihood ratio of probability theory. The length of logarithmic intervals can be computed in terms of this ratio by simple inspection since from (17)

$$\dot{x}_\tau(\theta) = \frac{\dot{f}_\tau(\theta)}{f_\tau(\theta)} = \log\left(\frac{f_1(\theta)}{f_0(\theta)}\right),$$

is independent of τ. Therefore the following statement holds.

Proposition 7: The length of the logarithmic path connecting two power spectral densities f_0 and f_1, is given by

$$\ell_{\text{logpath}} := \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\log\frac{f_1(\theta)}{f_0(\theta)}\right)^2 d\theta - \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_1(\theta)}{f_0(\theta)}d\theta\right)^2} \tag{20}$$

Proof: The proof follows in the arguments leading to the proposition. ∎

VII. Degradation of the smoothing error variance

In a way completely analogous to the previous sections we may consider the increase in the variance of the smoothing error when a wrong choice between two alternatives is used to identify a candidate smoothing filter.

Thus, we begin with two density functions f_1, f_2 and assume that $f_1^{-1}, f_2^{-1} \in L_1[-\pi, \pi)$. 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

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

Interestingly, this can be rewritten as follows

$$
\begin{aligned}
\rho_{\text{smooth}}(f_1, f_2) \; &= \; \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right)^2 f_1(\theta)^{-1} d\theta \right) \\
&\quad \times \frac{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)^{-1} d\theta \right)}{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right) f_1(\theta)^{-1} d\theta \right)^2} \\
&= \; \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right)^2 d\phi_1(\theta) \right) \\
&\quad \times \frac{1}{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right) d\phi_1(\theta) \right)^2} \\
&= \; \left(\frac{\sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right)^2 d\phi_1(\theta)}}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right) d\phi_1(\theta)} \right)^2
\end{aligned}
\tag{21}
$$

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 (21) shows the degradation as 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. However, the expression is homogeneous and does not depend on scaling of either of the two arguments f_1 or f_2.

Accordingly, we may define as a distance measure

$$
\begin{aligned}
\delta_{\text{smooth}}(f_1, f_2) \; &= \; \log(\rho_{\text{smooth}}(f_1, f_2)) \\
&= \; \log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right)^2 d\phi_1(\theta) \right) - \log \left(\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_1(\theta)}{f_2(\theta)} \right) d\phi_1(\theta) \right)^2 \right)
\end{aligned}
\tag{22}
$$

The presence of a data-dependent integration measure may by compared to the (normalized) Kullback-Leibler divergence in (11).

VIII. Reappraisal and generalizations

The expressions derived in the previous sections suggest that generalized means of the "likelihood"-like ratio $\Lambda := f_1/f_2$ and their logarithms may be used as distance measures between "shapes" of density functions f_1 and f_2. More specifically, we know that for any positive function Λ,

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

where $M_r(\Lambda)$ denotes the r-th generalized mean

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

Then

$$\delta_{r,s}(\Lambda) := \log(M_r(\Lambda)) - \log(M_s(\Lambda)) \geq 0$$

with a value which depends on how "far" Λ is from being constant. Hence, we may use $\delta_{r,s}(f_1/f_2)$ to quantify the distance between the "shapes" of f_1 and f_2, and since

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

is the geometric mean of Λ (see e.g., [3, page 23]), both δ, $\delta_{a/g}$ that we encountered earlier are special cases of the above. Weighted versions of weighted means may also be used for the same purpose, as suggested in Section VII. Lengths of geodesics as suggested in Section VI present another possibility. Indeed, a "zoo" of possible options emerges. Assessing practical and theoretical merits of each is the subject of a future project.

References

[1] S.M. Ali and S.D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Royal Stat. Soc.*, **28**: 131-142, 1966.

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

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

[4] L.M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. and Math. Phys.* vol. 7, pp. 200-217, 1967.

[5] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The Annals of Probability*, **19(4)**: 2032-2066, 1991.

[6] C. Byrnes, T.T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. on Signal Processing*, **48(11):** 3189-3206, November 2000.

[7] T.T. Georgiou, "Solution of the general moment problem via a one-parameter imbedding," *IEEE Trans. on Automatic Control*, **50(6):** 811-826, June 2005.

[8] T.T. Georgiou, "Relative Entropy and the multi-variable multi-dimensional Moment Problem," *IEEE Trans. on Information Theory*, **52(3):** 1052 - 1066, March 2006.

[9] T.T. Georgiou, "The maximum entropy ansatz in the absence of a time arrow: fractional-pole models," preprint, 18 pages: http://arxiv.org/abs/math/0601648/

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

[11] S. Haykin, **Nonlinear Methods of Spectral Analysis,** Springer-Verlag, New York, 247 pages, 1979.

[12] T.T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. on Information Theory*, **49(11)**, November 2003.

[13] 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.

[14] P. Masani, Recent trends in multivariate prediction theory, in **Multivariate Analysis**, P.R. Krishnaiah, Ed., Academic Press, pp. 351-382, 1966.

[15] W. Rudin, **Real and Complex Analysis**, 3rd edition, McGraw Hill, 1987.

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

[17] G. Szegö, "Über die randwerten eiher analytischen functionen," *Math. Ann.*, **84:** 232-244, 1921.

[18] S.R.S. Varadhan, **Probability Theory**, AMS, 2000.