An Intrinsic Metric for Power Spectral Density Functions

Tryphon T. Georgiou

***Abstract*—We present an intrinsic metric that quantifies distances between power spectral density functions. The metric was derived by Georgiou as the geodesic distance between spectral density functions with respect to a particular pseudo-Riemannian metric motivated by a quadratic prediction problem. We provide an independent verification of the metric inequality and discuss certain key properties of the induced topology.**

***Index Terms*—Information geometry, intrinsic metric, power spectral density functions.**

I. Metric Property

THIS letter builds on a recent report [4] where the present author introduced a natural pseudo-Riemannian metric on power spectral density functions of discrete-time stochastic processes, characterized geodesics, and computed geodesic distances. The *geodesic distance* between two power spectral density functions $f_i(\theta)$, with $i = 1, 2$, and $\theta \in [-\pi, \pi]$, was shown to be

$$d_g(f_1, f_2) := \sqrt{\int_{-\pi}^{\pi}\left(\log\frac{f_1(\theta)}{f_2(\theta)}\right)^2\frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi}\log\frac{f_1(\theta)}{f_2(\theta)}\frac{d\theta}{2\pi}\right)^2}. \quad (1)$$

In the following, we will provide a direct verification that $d_g(\cdot, \cdot)$ provides a pseudo-metric on the cone of power spectral density functions

$$\mathcal{D} := \{f : f(\theta) \geq 0 \text{ for } \theta \in [-\pi, \pi], f \in L_1[-\pi, \pi]\}.$$

As usual, L_1, L_2 denote Lebeague spaces of integrable and square-integrable functions, respectively.

The only reason $d_g(\cdot, \cdot)$ is a pseudo-metric and not a metric is because it is insensitive to scaling, i.e., $d_g(f_1, f_2) = d_g(f_1, \kappa f_2)$ for any $\kappa > 0$. Thus, it does not differentiate between spectral densities which only differ by a constant nonzero positive factor. Families of spectral density functions related in this way are referred to as spectral rays and form a set

$$\mathcal{R} := \{f \text{mod} \mathbb{R}_+ : f(\theta) \geq 0 \text{ for } \theta \in [-\pi, \pi], f \in L_1[-\pi, \pi]\} \quad (2)$$

of equivalence classes. Thus, $d_g(\cdot, \cdot)$ can be used to evaluate distances on $\mathcal{R}$ via comparing any two representatives on any two given spectral rays. In fact, $d_g(\cdot, \cdot)$ defines a metric on $\mathcal{R}$. It can be readily modified to provide a metric on $\mathcal{D}$, as well if, for instance, $|\int_{-\pi}^{\pi}(f_1(\theta) - f_2(\theta))\frac{d\theta}{2\pi}|$ or the absolute difference of any other generalized means is added on to differentiate the effect of scaling.

Before we proceed, we clarify how to evaluate $d_g(\cdot, \cdot)$ on all spectra in $\mathcal{D}$, including those that may vanish on a subset of the frequency interval $[-\pi, \pi]$ rendering $\log(f_1/f_2)^2$ nonintegrable. Clearly, when neither argument of $d_g(f_1, f_2)$ vanishes and f_i stays away from zero on $\in [-\pi, \pi]$ for $i = 1, 2$, then $\log f_i \in L_2[-\pi, \pi]$ and $d_g(f_1, f_2)$ is well defined and finite. However, if either f_i for $i \in \{1, 2\}$ vanishes on $[-\pi, \pi]$, then the integrals may diverge. However, since the root mean square of *any* function and, hence, of $\log(f_1/f_2)$, in particular, is always greater than or equal to its arithmetic mean (e.g., see [2]) it follows that:

$$\sqrt{\int_{-\pi}^{\pi}\left(\log\frac{f_1(\theta)}{f_2(\theta)}\right)^2\frac{d\theta}{2\pi}} \geq \int_{-\pi}^{\pi}\left(\log\frac{f_1(\theta)}{f_2(\theta)}\right)\frac{d\theta}{2\pi}. \quad (3)$$

Therefore, (1) gives either a nonnegative real value or has to be assigned the value $+\infty$. In conclusion, we complete the definition of $d_g(\cdot, \cdot)$ as follows. If

$$\log\frac{f_1}{f_2} \in L_2[-\pi, \pi] \quad (4)$$

in which case the left hand side of (3) is finite, $d_g(f_1, f_2)$ is evaluated using (1). If, however, (4) fails then, for consistency with (1), we assign

$$d_g(f_1, f_2) := \infty. \quad (5)$$

Clearly, failure of (4) can always be traced to at least one of $f_i (i \in \{1, 2\})$ failing to satisfy $\log f_i \in L_2[-\pi, \pi]$ (otherwise, necessarily, $\log(f_1/f_2) = (\log f_1 - \log f_2) \in L_2[-\pi, \pi]$).

Theorem 1: $d_g(\cdot, \cdot)$ defines a pseudo-metric on $\mathcal{D}$ and a metric on $\mathcal{R}$.

Proof: By definition, $d(\cdot, \cdot) \in [0, \infty) \cup \{+\infty\}$. It is also easy to observe that

$$d_g(f_1, f_2) = d_g(f_2, f_1). \quad (6)$$

To see this, note that $\log(f_1/f_2) = -\log(f_2/f_1)$ and that (1) is impervious to a sign change in front of the logarithms. Also, in the case one of $\log(f_1/f_2), \log(f_2/f_1)$ fails to be in L_2, so does the other, and again $d_g(f_1, f_2) = d_g(f_2, f_1)$ (both being ∞). Thus, (6) holds.

When $d_g(f_1, f_2) = 0$, the root mean square of the function $\log(f_1/f_2)$ is equal to its arithmetic mean, and this only happens (see [2]) when the function is constant. Indeed, $d_g(f_1, f_2) = 0$ implies

$$\begin{aligned}\log\frac{f_1(\theta)}{f_2(\theta)} &= c \in \mathbb{R} \text{ for all } \theta \in [-\pi, \pi]\\ &\Rightarrow f_1 = \kappa f_2\\ &\Rightarrow f_1 \text{mod}\mathbb{R} = f_2 \text{mod}\mathbb{R}\end{aligned}$$

Manuscript received August 19, 2006; revised November 10, 2006. This work was supported by the National Science Foundation, the Air Force Office of Scientific Research, and the Vincentinne Hemes–Luh Chair. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Deniz Erdogmus.

The author is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA.

Digital Object Identifier 10.1109/LSP.2006.891315

since $\kappa = e^c$ is a constant. Thus, $d_g(\cdot,\cdot)$ separates the elements of $\mathcal{R}$.

We finally establish the triangular inequality. So, let us consider $f_i \in \mathcal{D}$ for $i \in \{1,2,3\}$. We will show that

$$d_g(f_1, f_2) + d_g(f_2, f_3) \geq d_g(f_1, f_3). \quad (7)$$

We first argue the case when $d_g(f_1, f_3) = \infty$. It suffices to show that one of the left hand side terms is also infinity. Assume the contrary, i.e., that

$$\log \frac{f_1}{f_2} \in L_2[-\pi,\pi] \text{ as well as } \log \frac{f_2}{f_3} \in L_2[-\pi,\pi].$$

It readily follows that $\log(f_1/f_3) = \log(f_1/f_2) + \log(f_2/f_3) \in L_2[-\pi,\pi]$ which contradicts the assumption that $d_g(f_1, f_3) = \infty$. Thus, at least one of $d_g(f_1, f_2), d_g(f_2, f_3)$ is infinity and the triangular inequality holds. Of course, if $\log(f_1/f_3)$ is finite and any of $d_g(f_1, f_2), d_g(f_2, f_3)$ takes the value ∞, the triangular inequality holds anyway.

We now argue the case when all three $d_g(f_1, f_2), d_g(f_2, f_3)$ and $d_g(f_1, f_3)$ are finite. To this end we square both sides of (7) and utilize

$$\log \frac{f_1}{f_3} = \log \frac{f_1}{f_2} + \log \frac{f_2}{f_3} \quad (8)$$

to simplify the resulting expression and deduce the following inequality:

$$\begin{aligned} &\sqrt{\int_{-\pi}^{\pi} \left(\log \frac{f_1}{f_2}\right)^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \log \frac{f_1}{f_2} \frac{d\theta}{2\pi}\right)^2} \\ &\quad \times \sqrt{\int_{-\pi}^{\pi} \left(\log \frac{f_2}{f_3}\right)^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \log \frac{f_2}{f_3} \frac{d\theta}{2\pi}\right)^2} \\ &\quad \geq \int_{-\pi}^{\pi} \left(\log \frac{f_1}{f_2} \log \frac{f_2}{f_3}\right) \frac{d\theta}{2\pi} \\ &\quad - \left(\int \log \frac{f_1}{f_2} \frac{d\theta}{2\pi}\right) \left(\int \log \frac{f_2}{f_3} \frac{d\theta}{2\pi}\right). \end{aligned} \quad (9)$$

Thus, the two inequalities (9) and (7) are equivalent to one another, and, therefore, in order to ascertain (7), it suffices to establish (9).

To this end, let $\alpha := \log(f_1/f_2)$, $\beta := \log(f_2/f_3)$ and rewrite (9) in the form

$$\begin{aligned} &\sqrt{\int_{-\pi}^{\pi} \alpha^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \alpha \frac{d\theta}{2\pi}\right)^2} \sqrt{\int_{-\pi}^{\pi} \beta^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \beta \frac{d\theta}{2\pi}\right)^2} \\ &\qquad \geq \int_{-\pi}^{\pi} (\alpha\beta) \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \alpha \frac{d\theta}{2\pi}\right) \left(\int_{-\pi}^{\pi} \beta \frac{d\theta}{2\pi}\right). \end{aligned} \quad (10)$$

Since (10) is homogeneous in both α and β, scaling of either leaves it unaffected. Therefore, if

$$\sigma_\alpha = \sqrt{\int_{-\pi}^{\pi} \alpha^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \alpha \frac{d\theta}{2\pi}\right)^2}$$
$$\sigma_\beta = \sqrt{\int_{-\pi}^{\pi} \beta^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \beta \frac{d\theta}{2\pi}\right)^2}$$

and $a = \alpha/\sigma_\alpha, b = \beta/\sigma_\beta$, the inequality (10) is equivalent to

$$1 \geq \int_{-\pi}^{\pi} (ab) \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} a \frac{d\theta}{2\pi}\right) \left(\int_{-\pi}^{\pi} b \frac{d\theta}{2\pi}\right) \quad (11)$$

with the side conditions

$$\int_{-\pi}^{\pi} a^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} a \frac{d\theta}{2\pi}\right)^2 = 1 \quad (12)$$
$$\int_{-\pi}^{\pi} b^2 \frac{d\theta}{2\pi} - (\int_{-\pi}^{\pi} b \frac{d\theta}{2\pi})^2 = 1. \quad (13)$$

However, the validity of (11) follows trivially from the standard inequality

$$\int_{-\pi}^{\pi} (a-b)^2 \frac{d\theta}{2\pi} \geq \left(\int_{-\pi}^{\pi} (a-b) \frac{d\theta}{2\pi}\right)^2$$

after we expand the squares on both sides and use (12) and (13) to simplify the resulting expressions. Thus, (11) with (12)–(13) holds $\Rightarrow$ (10) $\Rightarrow$ (9) $\Rightarrow$ (7), and this completes the proof. ∎

Remark: The definition of $d_g(\cdot,\cdot)$ distinguishes two classes of power spectral densities according to whether their logarithm is square integrable or not. The first class, $\mathcal{D}_{\text{interior}} := \{f \in \mathcal{D} : \log f \in L_2[-\pi,\pi]\}$, can be thought of as "interior" points lying to within a finite distance from one another, and to within a finite distance from constant nonzero power spectral densities. The second class, with logarithms that fail to be square integrable, contains power spectral densities which lie at an infinite distance from any density in $\mathcal{D}_{\text{interior}}$. On the other hand, power spectral densities are traditionally differentiated according to whether the underlying process is deterministic or not. More specifically, a stochastic process is said to be *nondeterministic* (in the sense of Kolmogoroff) if the variance of the one-step-ahead prediction error cannot be made arbitrarily small. In turn, this property is characterized by the log-integrability of the corresponding power spectral density function (see [6] and [7]), i.e.,

$$\mathcal{D}_{\text{nondeterministic}} := \{f \in \mathcal{D} : \log f \in L_1[-\pi,\pi]\}.$$

Thus, it is interesting to observe that

$$\mathcal{D}_{\text{interior}} \subset \mathcal{D}_{\text{nondeterministic}}$$

and that finite neighborhoods in $\mathcal{D}_{\text{interior}}$ contain nondeterministic power spectra *only*.

II. Riemannian Geometry, Geodesics, and Intrinsic Metrics

We explain the geometric significance of $d_g(\cdot,\cdot)$ recapitulating some of the development in [4].

The starting point is a prediction problem and the degradation of the variance of the prediction error when the choice of predictor is based on the wrong choice among two alternatives. More specifically, let f_1, f_2 represent spectral density functions of discrete-time zero-mean stochastic processes $u_{f_i}(k)$ ($i \in \{1,2\}$ and $k \in \mathbb{Z}$), and let $p_{f_i}(\ell)$ ($\ell \in \{1,2,3,\ldots\}$) represent values for the coefficients that minimize the linear prediction error variance

$$\mathcal{E}\left\{|u_{f_i}(0) - \sum_{\ell=1}^{\infty} p(\ell)u_{f_i}(-\ell)|^2\right\}.$$

Thus, the optimal set of coefficients depends on the power spectral density function of the process, a fact which is duly acknowledged by the subscript in the notation $p_{f_i}(\ell)$. Here, as usual, $\mathcal{E}\{\,\cdot\,\}$ denotes the expectation operator. It is reasonable to consider as a distance between f_1 and f_2 the degradation of predictive error variance when the coefficients $p(\ell)$ are selected assuming one of the two, and then used to predict a stochastic process corresponding to the other spectral density function.

The ratio of the "degraded" predictive error variance over the optimal error variance

$$\rho(f_1, f_2) := \frac{\mathcal{E}\left\{u_{f_1}(0) - \sum_{\ell=1}^{\infty} p_{f_2}(\ell)u_{f_1}(-\ell)|^2\right\}}{\mathcal{E}\left\{u_{f_1}(0) - \sum_{\ell=1}^{\infty} p_{f_1}(\ell)u_{f_1}(-\ell)|^2\right\}}$$

turns out to be equal to the ratio of the *arithmetic* over the *geometric* means of the fraction of the two spectral density functions, namely (see [4])

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

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

$$\gamma(f_1, f_2) := \rho(f_1, f_2) - 1 \in [0, \infty]$$

quantifies the dissimilarity between the "shapes" of the two spectral densities. The same applies to $\log(\rho(f_1, f_2))$. Both expressions are analogous to "divergences" of Information Theory and in fact, $\log(\rho(f_1, f_2))$ coincides with a "gain optimized" version of the Itakura-Saito distortion measure in the speech processing literature (cf. [5, p. 371, Section C]).

Considering the distance $\gamma(f, f + \Delta)$ between a nominal power spectral density f and a perturbations $f + \Delta$, and eliminating cubic terms and beyond, leads (modulo a scaling factor of 2) to the Riemannian pseudo-metric in $\mathcal{D}_{\text{interior}}$ which is given by the following quadratic differential form:

$$g_f(\Delta) := \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)}\right)^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \frac{\Delta(\theta)}{f(\theta)} \frac{d\theta}{2\pi}\right)^2. \quad (14)$$

Interestingly, geodesic paths f_τ ($\tau \in [0, 1]$) connecting spectral densities f_0, f_1 and having minimal length

$$\int_0^1 \sqrt{g_{f_\tau}\left(\frac{\partial f_\tau}{\partial \tau}\right)} d\tau$$

can be explicitely characterized [4]. They turn out to be logarithmic intervals

$$f_\tau(\theta) = f_0^{1-\tau}(\theta) f_1^{\tau}(\theta) \text{ for } \tau \in [0, 1] \quad (15)$$

between the two extreme points. Furthermore, the length along such geodesics is precisely $d_g(f_0, f_1)$ as given in (1).

The closed form of the geodesic path allows us to verify directly that any two power spectral densities at a finite distance from one another, can be connected with a path of the same length. A topological space with such a property is said to be a *length-space* and the metric is said to be *intrinsic*. The fact that $d_g(\cdot, \cdot)$ is intrinsic can be readily verified and this is done below.

Proposition 2: $d_g(\cdot, \cdot)$ is intrinsic on $\mathcal{D}$ and $\mathcal{R}$.

Proof: By direct substitution into (1), we can verify that for any f_0, f_1 such that $d_g(f_0, f_1) < \infty$, any $\tau \in [0, 1]$, and with f_τ defined as in (15), $d_g(f_0, f_\tau) = \tau d_g(f_0, f_1), d_g(f_\tau, f_1) = (1 - \tau) d_g(f_0, f_1)$, and even $d_g(f_\tau, f_{\tau+d\tau}) = d_g(f_0, f_1) d\tau$. It readily follows that the length of the path $\int_0^1 d_g(f_\tau, f_{\tau+d\tau}) = \int_0^1 d_g(f_0, f_1) d\tau$ equals the distance $d_g(f_0, f_1)$ between the end points. ∎

III. Concluding Thoughts

The logarithmic map $f \mapsto \log f$ takes the geodesics in (15) into straight lines inside a linear space. Thus, the L_2-norm of the difference between logarithms of power spectral densities, i.e., $|\log(f_1/f_2)|_2$, presents an attractive alternative as a convenient metric with properties similar to those of $d_g(\cdot, \cdot)$.

It has been pointed out by an anonymous referee that formulae involving the Euclidean distance between logarithms of eigenvalues of covariance pencils of multivariable Gaussian distributions [1, p.236, item (9)] may point to discrete analogs of (14).

Finally, it is interesting to compare the differential structure of g_f with an analogous differential structure introduced in Information Geometry [1] for propability densities. Indeed, if $f, f + \Delta$ represent probability densities on $[-\pi, \pi]$, then

$$\begin{aligned} g_{\text{Fisher},f}(\Delta) &= \int_{-\pi}^{\pi} \left(\frac{\Delta(\theta)}{f(\theta)}\right)^2 f(\theta) \frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi} \frac{\Delta(\theta)^2}{f(\theta)} \frac{d\theta}{2\pi} \end{aligned} \quad (16)$$

which is known as the *Fisher information metric*, is a natural Riemannian metric (cf. [1, p. 28] and [3]). This metric can be derived from the Kullback-Leibler divergence, in a way analogous to the derivation of g_f from $\gamma(\cdot, \cdot)$, via eliminating higher than quadratic terms. Direct comparison reveals that the powers of $f(\theta)$ in (14) and (16) are different. Thus, it is worth underscoring the fact that in either differential structure, geodesics and geodesic lengths can be computed explicitly. Yet, while $f \mapsto \log f$ maps geodesics (15) of g_f into straight lines, the $g_{\text{Fisher},f}$-manifold of probability densities is curved [3].

References

[1] S. Amari and H. Nagaoka, "Methods of information geometry," in *Translations of Mathematical Monographs*. Oxford, U.K.: Oxford Univ. Press, 2000, AMS.

[2] E. F. Beckenbach and R. Bellman, *Inequalities*. Berlin-Heidelberg, Germany: Springer-Verlag, 1965, p. 198.

[3] L. L. Campbell, "The relation between information theory and the differential geometric approach to statistics," *Inf. Sci.*, vol. 35, pp. 199–210, 1985.

[4] T. T. Georgiou, "Distances between power spectral densities," arXiv:math.OC/0607026, Jul. 2006.

[5] R. Gray, A. Buzo, A. Gray, and Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 367–376, Aug. 1980.

[6] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*. New York: Chelsea, 1958.

[7] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 2005.