

# Conal Distances Between Rational Spectral Densities

Giacomo Baggio, Augusto Ferrante, and Rodolphe Sepulchre

**Abstract**—The paper generalizes Thompson and Hilbert metric to the space of spectral densities. The resulting complete metric space has the differentiable structure of a Finsler manifold with explicit geodesics. The corresponding distances are filtering invariant, can be computed efficiently, and admit geodesic paths that preserve rationality; these are properties of fundamental importance in many engineering applications.

**Index Terms**—Rational spectral densities, conal distances, Finsler geometry, Thompson metric, Hilbert metric, linear filtering, spectral estimation, speech morphing.

## I. INTRODUCTION

In recent years the design and analysis of distances between spectral densities have received a renewed interest in the control and signal processing community (see [1], [2] for two recent surveys on this topic). This interest primarily stems from a large number of applications in which the problem of quantifying dissimilarities between spectral densities is of crucial importance, such as spectral estimation [3]–[10], speech processing [11]–[16], and time-series clustering [17]–[22], to cite a few.

The design of distances with the aim of solving computational engineering problems is a rich topic because of the interplay between mathematical, modelling, and computational considerations. Mathematical considerations include endowing the underlying space with a differential manifold structure such that the distance between two points corresponds to the length of a minimal geodesic. This is especially relevant when dealing with problems involving approximation, smoothing, and averaging of spectral densities, e.g., in the context of speech morphing [13], [14], [16]. The classical framework is Riemannian geometry, in which the differential structure involves an inner product. The present paper uses the broader framework of Finsler geometry, where the differential structure only requires a norm. Modelling considerations include endowing the distance with suitable invariance properties, such that the mathematical distance is consistent with what is modelled. In the context of spectral densities, filtering invariance emerges as a property that should hold when spectral densities model second-order stationary stochastic processes.

G. Baggio is with the Department of Mechanical Engineering, University of California at Riverside, Riverside, CA 92521 USA. E-mail: giacomob@ucr.edu. A. Ferrante is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo, 6/B I-35131 Padova, Italy. E-mail: augusto@dei.unipd.it. R. Sepulchre is with the Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK. E-mail: r.sepulchre@eng.cam.ac.uk.

The research leading to these results has received funding from the European Research Council under the Advanced ERC Grant Agreement Switchlet n.670645.

Invariance properties are receiving increasing attention in engineering, because they tend to make algorithms less sensitive to modelling assumptions. Computational considerations include the existence of an algorithmic framework to perform the calculations necessary to the considered engineering problem, starting with the evaluation of the distance itself. They are of primary importance in high dimensional problems, and, a fortiori, so for infinite dimensional objects such as spectral densities.

The starting point in this paper is to acknowledge that the space of spectral densities is a *cone* and to revisit two classical distances that have been studied in cones: the part metric (often called Thompson metric) and the projective metric introduced by Hilbert. Applying this classical framework to the space of *rational* spectral densities, which seems novel, we show that the resulting metrics have a number of particularly desirable properties:

- they are filtering invariant, a natural and desirable property in many applications;
- their calculations boil down to evaluating the  $\mathcal{H}_\infty$ -norm of minimum-phase spectral factors, a problem extensively studied in the control literature that can be performed via efficient routines;
- they endow the cone of spectral densities with a Finsler geometry featuring explicit geodesic paths that can be chosen to be *rational*.

In particular, we show that the Thompson metric is a close relative of a Riemannian metric recently studied in [23]. However, the latter does not enjoy all the above-listed properties.

*Paper structure.* After some preliminary definitions and necessary notations, in Section II, we illustrate some applicative scenarios that motivate the introduction of the new metrics. Section III reviews the Finsler geometry of cones, with a special emphasis on the cone of positive definite matrices. Section IV applies this geometry to the cone of rational spectral densities and discuss the properties inherited by this geometry. Section V presents an applicative example of spectral interpolation via geodesic paths of the introduced metric, in the context of speech processing. Lastly, Section VI collects some concluding remarks and future research directions.

*Notation.* As usual, we denote by  $\mathbb{R}$ ,  $\mathbb{C}$ ,  $\mathbb{R}^{n \times m}$ ,  $\mathbb{C}^{n \times m}$ , and  $\mathbb{R}^{n \times m}(z)$ , the set of real numbers, complex numbers,  $n \times m$  real matrices,  $n \times m$  complex matrices, and  $n \times m$  real matrix-valued rational functions in  $z \in \mathbb{C}$ , respectively.  $\mathbb{T}$  and  $\mathbb{D}$  will denote, respectively, the unit circle, and the open unit disk in the complex plane.  $\mathbb{R}_*^{n \times n}(z)$  will denote the set of  $n \times n$  real matrix-valued rational functions of full rank on  $\mathbb{T}$ ; it forms a multiplicative group. Given  $A \in \mathbb{C}^{n \times n}$ ,  $A^\top$ ,  $A^*$ ,

$\text{tr}(A)$ , and  $\|A\|_F$  stand for the transpose, Hermitian transpose, trace and Frobenius norm of  $A$ , respectively. Let  $\mathcal{L}_2^n[-\pi, \pi]$  be the space of  $n$ -dimensional vector-valued functions on  $\mathbb{T}$  that are square integrable w.r.t. the normalized Lebesgue measure. The space  $\mathcal{L}_2^n[-\pi, \pi]$  endowed with the inner product  $\langle f, g \rangle_2 := \int_{-\pi}^{\pi} f(e^{j\theta})^* g(e^{j\theta}) \frac{d\theta}{2\pi}$ ,  $f, g \in \mathcal{L}_2^n[-\pi, \pi]$ , forms an Hilbert space. We let  $\mathcal{S}_+^{n \times n}$  denote the cone of  $n \times n$  positive semi-definite Hermitian matrices and  $\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  the cone of  $n \times n$  bounded positive self-adjoint operators on  $\mathcal{L}_2^n[-\pi, \pi]$ , namely,

$$\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T}) := \{ \Phi: \mathbb{T} \rightarrow \mathbb{C}^{n \times n} : \Phi(e^{j\theta}) = \Phi(e^{j\theta})^*, \forall e^{j\theta} \in \mathbb{T}, \\ \text{and } \langle f, \Phi f \rangle_2 \geq 0, \forall f \in \mathcal{L}_2^n[-\pi, \pi] \}.$$

Henceforth, elements of  $\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  will be thought of as discrete-time *spectral densities*<sup>1</sup> and  $\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T}) \subset \mathcal{S}_+^{n \times n}(\mathbb{T})$  will denote the subset of  $n \times n$  real *rational* bounded discrete-time spectral densities. Given  $A(z) \in \mathbb{R}^{n \times m}(z)$ , we let  $A^*(z) := A^\top(1/z)$  and we use  $A^{-*}(z)$  as a shorthand for  $[A^*(z)]^{-1}$ . A rational matrix  $W(z) \in \mathbb{R}^{n \times r}(z)$  is called a spectral factor of the spectral density  $\Phi(z) \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$ , if it satisfies  $\Phi(z) = W(z)W^*(z)$ . If the spectral factor  $W(z)$  is analytic in (an open set containing) the complement of  $\mathbb{D}$  with (generalized) inverse analytic in the complement of the closure of  $\mathbb{D}$ , then  $W(z)$  is called minimum-phase (or canonical) spectral factor. The minimum-phase spectral factor of a rational spectral density  $\Phi(z) \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  always exists and is essentially unique, that is, is unique up to post-multiplication by orthogonal  $r \times r$  matrices, where  $r$  is the normal rank (that is, the a. e. rank on  $\mathbb{T}$ ) of  $\Phi$  [25]. To conclude, we recall that, given a  $n \times m$  matrix-valued function  $G: \mathbb{T} \rightarrow \mathbb{C}^{n \times m}$  that is (essentially) bounded on  $\mathbb{T}$ , the  $\mathcal{L}_\infty$ -norm of  $G$  is defined as  $\|G\|_{\mathcal{L}_\infty} := \text{ess sup}_{\theta \in [-\pi, \pi]} \sigma_{\max}(G(e^{j\theta}))$  where  $\sigma_{\max}(A)$  denotes the largest singular value of  $A \in \mathbb{C}^{n \times m}$ . If, furthermore,  $G$  can be analytically extended in an open set that contains the complement of  $\mathbb{D}$ , then its  $\mathcal{L}_\infty$ -norm coincides with the  $\mathcal{H}_\infty$ -norm  $\|G\|_{\mathcal{H}_\infty} := \sup_{z \in \mathbb{C}: |z| > 1} \sigma_{\max}(G(z))$ .

## II. MOTIVATIONAL EXAMPLES

As mentioned in the introduction, engineering applications motivate the design of distances that (i) are computable, (ii) possess a differential structure amenable to interpolation and extrapolation, and (iii) are invariant to pre-processing of the data. In what follows, we illustrate those concrete motivations in representative specific applications.

*Example 1 (Time-series clustering [17]–[19]):* Let  $\{y_i(k)\}_{k \in M_i}$ ,  $M_i \subset \mathbb{Z}$ ,  $i = 1, 2, \dots, N$ , be a set of  $p$ -dimensional time series data collected from measurements and representing the (noisy) behavior of some physical or engineering systems. Each time series is modelled by a dynamical model estimated from the raw data. The estimated model is usually taken to be linear and time-invariant, and admits the (formal) input-output representation

$$y_i(k) = W_i(z)e_i(k), \quad i = 1, 2, \dots, N, \quad (1)$$

<sup>1</sup>We remark that the standard definition of spectral density needs integrability only, see, e.g., [24]. In this paper, however, we restrict the attention to a subset of “well-behaved” spectral densities that includes the (particularly important) class of bounded rational spectral densities.

where  $W_i(z)$  is a  $p \times m$  rational matrix and  $\{e_i(k)\}_{k \in \mathbb{Z}}$  an  $m$ -dimensional white noise process. In many practical applications ranging from econometrics to biology, one needs to classify the different times series data  $\{y_i(k)\}_{k \in M_i}$  in different groups, based on their similarity. This procedure is commonly known as *time-series clustering*.

The desired clustering can be performed by defining a suitable distance index between corresponding linear models in (1). Since the spectral density of the  $i$ -th time-series is  $\Phi_i(z) = W_i(z)W_i^*(z)$ , the distance can be defined in the space of (rational) spectral densities. In many real scenarios, the amount of time series data is massive and one seeks a distance that is, at the same time, computationally tractable and accurate.

*Example 2 (“Geodesic” speech morphing [13], [16]):* Consider two digital speech sources coming from two different individuals, say A and B, and described by time series  $\{y_A(k)\}_{k=1}^N$  and  $\{y_B(k)\}_{k=1}^N$ , respectively. The *morphing* of these two audio signal consists of gradually deforming the speech signal of A into the speech signal of B, creating a new hybrid signal that should preserve the speech-like quality and content of the original sources. Nowadays, there are many applications that benefit from such algorithms, especially in the areas of multimedia engineering and entertainment.

By segmenting the two time series into  $M < N$  approximately stationary fragments, one can first estimate the spectral density of each fragment, say  $\{\Phi_{A,i}\}_{i=1}^M$  and  $\{\Phi_{B,i}\}_{i=1}^M$ , and then perform morphing using a *path* connecting the spectral densities of each fragment. A *geodesic* between  $\Phi_{A,i}$  and  $\Phi_{B,i}$  is a particularly convenient and natural choice for such a path. Of course, this requires the definition of a suitable metric in the space of spectral densities that equips the latter space with a differential structure. Thus, a geodesic interpolation path between  $\Phi_{A,i}$  and  $\Phi_{B,i}$  provides a geometric solution to the morphing engineering question.

*Example 3 (THREE-like spectral estimation [3]–[7]):* Let  $\{y(t)\}_{t \in \mathbb{Z}}$  be an  $n$ -dimensional zero-mean stationary process and let  $\Psi \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  be an a priori estimate of the unknown spectral density of this process. Consider the bank of filters described by the transfer function

$$G(z) = (zI - A)^{-1}B, \quad A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m},$$

with  $A$  strictly Schur stable,  $B$  of full column rank, and the pair  $(A, B)$  reachable. We assume to have an estimate of the asymptotic state covariance  $\Sigma \in \mathcal{S}_+^{n \times n}$  of the system with transfer function  $G(z)$  and input the process  $\{y(t)\}_{t \in \mathbb{Z}}$ . The task is to estimate the spectral density of  $\{y(t)\}_{t \in \mathbb{Z}}$  based on the available information. Typically, the prior  $\Psi$  is not consistent with the state covariance  $\Sigma$ . Therefore, it is necessary to find a spectral density which is as closest as possible, in some suitable sense, to  $\Psi$ , and, additionally, satisfies the “consistency” condition

$$\int_{-\pi}^{\pi} G(e^{j\theta})\Phi(e^{j\theta})G^*(e^{j\theta}) \frac{d\theta}{2\pi} = \Sigma.$$

This formulation leads to the following constrained optimization problem

$$\begin{aligned} \min_{\Phi \in \mathcal{S}_+^{m \times m}(\mathbb{T})} d(\Psi, \Phi) \\ \text{s.t. } \int_{-\pi}^{\pi} G(e^{j\theta}) \Phi(e^{j\theta}) G^*(e^{j\theta}) \frac{d\theta}{2\pi} = \Sigma \end{aligned}$$

where  $d: \mathcal{S}_+^{m \times m}(\mathbb{T}) \times \mathcal{S}_+^{m \times m}(\mathbb{T}) \rightarrow [0, \infty)$  is a suitable distance function in the cone of spectral densities.

One crucial aspect in the above estimation problem concerns the choice of the distance measure  $d(\cdot, \cdot)$  to minimize. In order to be effective, this distance should satisfy some properties that naturally arise from the estimation setting. For instance, one natural requirement is that the distance between the prior process (described by spectral density  $\Psi$ ) and the unknown process must be left unchanged if we filter both processes via the same filter. Another often desired property is that the distance is projective meaning that it is unaffected by scalings (in this case the “shape” of the unknown spectrum is actually estimated). Finally, the distance must be amenable to algorithmic optimization, which benefits from properties such as convexity and requires efficient numerical estimation of the distance and its first or second derivatives.

### III. DISTANCES IN CONES

Let  $\mathcal{K}$  be a closed, solid, pointed, convex cone defined in a real Banach space  $\mathcal{B}$  with norm  $\|\cdot\|_{\mathcal{B}}$ , that is, a closed subset  $\mathcal{K}$  with the properties that: (i) the interior of  $\mathcal{K}$ , denoted by  $\mathring{\mathcal{K}}$ , is non-empty, (ii)  $\mathcal{K} + \mathcal{K} \subseteq \mathcal{K}$ , (iii)  $\mathcal{K} \cap -\mathcal{K} = \{0\}$ , (iv)  $\lambda \mathcal{K} \subseteq \mathcal{K}$  for all  $\lambda \geq 0$ . The cone  $\mathcal{K}$  induces a partial ordering  $\leq_{\mathcal{K}}$  on  $\mathcal{B}$  by

$$x \leq_{\mathcal{K}} y \iff y - x \in \mathcal{K}.$$

For  $x, y \in \mathcal{K}$ , we say that  $y$  dominates  $x$  if there exists  $\beta > 0$  such that  $x \leq_{\mathcal{K}} \beta y$ . We write  $x \sim_{\mathcal{K}} y$  if  $y$  dominates  $x$ , and  $x$  dominates  $y$ . The relation  $\sim_{\mathcal{K}}$  is an equivalence relation on  $\mathcal{K}$ . The corresponding equivalence classes are called *parts* or *components* of  $\mathcal{K}$ .

Given two elements  $x, y \in \mathcal{K} \setminus \{0\}$ , we define the following quantities

$$M(x, y) := \inf\{\lambda : x \leq_{\mathcal{K}} \lambda y\}, \quad (2)$$

if the set is non-empty, and  $M(x, y) := \infty$  otherwise, and

$$m(x, y) := \sup\{\mu : \mu y \leq_{\mathcal{K}} x\} = \frac{1}{M(y, x)}. \quad (3)$$

*Definition 1 ([26], [27]):* The *Hilbert (projective) metric* and the *Thompson (part) metric* between elements  $x, y \in \mathcal{K} \setminus \{0\}$  are defined respectively by

$$d_H(x, y) := \log \frac{M(x, y)}{m(x, y)}, \quad (4)$$

$$d_T(x, y) := \log \max\{M(x, y), M(y, x)\}, \quad (5)$$

if  $x \sim_{\mathcal{K}} y$ , and  $d_H(x, y) = d_T(x, y) := \infty$ , otherwise.

As a simple example, consider  $\mathcal{B} = \mathbb{R}^n$  and  $\mathcal{K}$  to be the positive orthant of  $\mathbb{R}^n$ , i.e.  $\mathcal{K} := \{(x_1, \dots, x_n) : x_i \geq 0, 1 \leq i \leq n\}$ . In this case, for  $x, y \in \mathring{\mathcal{K}}$ , it holds

$$M(x, y) = \max_i \{x_i / y_i\},$$

$$m(x, y) = \min_i \{x_i / y_i\},$$

so that Hilbert and Thompson metrics on  $\mathring{\mathcal{K}}$  read, respectively, as

$$d_H(x, y) = \log \frac{\max_i \{x_i / y_i\}}{\min_i \{x_i / y_i\}},$$

$$d_T(x, y) = \log \max \left\{ \max_i \{x_i / y_i\}, \max_i \{y_i / x_i\} \right\}.$$

Thompson metric is a bona fide distance<sup>2</sup> on each part of the cone  $\mathcal{K}$  (and, in particular, on the interior  $\mathring{\mathcal{K}}$ ). Each part of  $\mathcal{K}$  is a *complete* metric space with respect to this metric provided that  $\mathcal{K}$  is normal, i.e., there exists  $\gamma > 0$  such that  $\|x\|_{\mathcal{B}} \leq \gamma \|y\|_{\mathcal{B}}$  holds whenever  $0 \leq_{\mathcal{K}} x \leq_{\mathcal{K}} y$  [27]. Hilbert metric is a distance between *rays* in each part of  $\mathcal{K}$ :  $d_H(x, y) = 0$ ,  $x, y \in \mathcal{K}$ ,  $x \sim_{\mathcal{K}} y$ , if and only if  $x = \lambda y$  with  $\lambda > 0$ .

Hilbert and Thompson metric have been of great interest to analysts, especially for their contractivity properties. As a matter of fact, many naturally occurring maps in analysis, both linear and non-linear, are either non-expansive or contractive with respect to these metrics [26], [28], [29]. Moreover, it has been proven that among all projective distances  $d$  on  $\mathcal{K}$  for which the positive linear transformations are contractive w.r.t.  $d$ , Hilbert metric is the one with the best possible contraction ratio [30].

Thompson and Hilbert metric endow the cone with a structure of Finsler manifold [31]. In the finite-dimensional case, the interior of the cone  $\mathcal{K}$  defines an  $n$ -dimensional manifold and the tangent space at each point may be identified with  $\mathbb{R}^n$ . Defining the norm

$$\|v\|_x^T := \inf\{\alpha > 0 : -\alpha x \leq_{\mathcal{K}} v \leq_{\mathcal{K}} \alpha x\} \quad (6)$$

on the tangent space at each point  $x \in \mathring{\mathcal{K}}$ , the length of any differentiable curve  $\gamma: [a, b] \rightarrow \mathring{\mathcal{K}}$  is defined as

$$\ell(\gamma) := \int_a^b \|\gamma'(t)\|_{\gamma(t)}^T dt.$$

Thompson distance between any two points is recovered by minimizing over all paths connecting the points, namely

$$d_T(x, y) = \inf\{\ell(\gamma) : \gamma \in C^1[x, y]\},$$

where  $C^1[x, y]$  denotes the set of all differentiable paths  $\gamma: [a, b] \rightarrow \mathring{\mathcal{K}}$  such that  $\gamma(0) = x$  and  $\gamma(1) = y$ . Hilbert metric is obtained along the same lines when the norm above is replaced by the semi-norm

$$\|v\|_x^H := M(v, x) - m(v, x).$$

The Finslerian nature of Hilbert and Thompson geometries allows for the definition of minimal geodesics connecting two points in the interior of the cone  $\mathcal{K}$ . Differently from the Riemannian framework, minimal geodesics connecting two points are usually not unique [29, Ch. 2]. An explicit class

<sup>2</sup>We recall that a bona fide metric or distance function on a set  $X$  is a function  $d: X \times X \rightarrow [0, \infty)$  satisfying the following conditions for all  $x, y, z \in X$ : (i)  $d(x, y) \geq 0$ , (ii)  $d(x, y) = 0$  if and only if  $x = y$ , (iii)  $d(x, y) = d(y, x)$ , (iv)  $d(x, z) \leq d(x, y) + d(y, z)$ .

of minimal geodesics for the Thompson metric connecting  $x, y \in \hat{\mathcal{K}}$ , is given by,  $\chi: [0, 1] \rightarrow \hat{\mathcal{K}}$ ,

$$\chi(\tau) = \begin{cases} \left( \frac{\beta^\tau - \alpha^\tau}{\beta - \alpha} \right) y + \left( \frac{\beta\alpha^\tau - \alpha\beta^\tau}{\beta - \alpha} \right) x, & \text{if } \beta \neq \alpha, \\ \alpha^\tau x, & \text{if } \beta = \alpha, \end{cases} \quad (7)$$

where  $\beta := M(y, x)$  and  $\alpha := m(y, x)$ . This geodesic path defines a “projective” straight line in the cone [32].

Such geodesic paths are not unique. For instance, a distinct type of geodesic paths connecting two positive definite matrices  $X$  and  $Y$  is given by

$$\varphi_T(\tau) = X^{1/2} (X^{-1/2} Y X^{-1/2})^\tau X^{1/2}, \quad \tau \in [0, 1]. \quad (8)$$

This path is in fact the (unique, up to a re-parametrization) Riemannian geodesic of  $\mathcal{S}_+^{n \times n}$  connecting  $X$  to  $Y$  with respect to the affine invariant metric, see e.g. [33, Thm. 6.1.6]. The corresponding geodesic path in (8) w.r.t. Hilbert metric reads [29, Prop. 2.6.8]

$$\varphi_H(\tau) = \frac{\varphi_T(\tau)}{\text{tr}(\varphi_T(\tau))}, \quad \tau \in [0, 1], \quad (9)$$

where the latter path connects two unit-trace elements  $X, Y \in \hat{\mathcal{S}}_+^{n \times n}$  which are the representatives of the corresponding projective rays  $\mu X, \mu Y, \mu > 0$ , respectively.

Finally, we remark that the Finslerian framework so far discussed for the case of finite-dimensional spaces applies without any substantial change to the case of infinite-dimensional manifolds of bounded positive self-adjoint operators on a Hilbert space. For further details on this extension we refer to the works by Corach and co-workers [34], [35], and, in particular, to [36], [37].

#### IV. FINSLERIAN DISTANCES IN $\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$

Since spectral densities can be thought of as bounded positive self-adjoint operators on the Hilbert space  $\mathcal{L}_2^n[-\pi, \pi]$ , the framework outlined in the previous section provides Finslerian distances in the cone  $\mathcal{S}_+^{n \times n}(\mathbb{T})$ , and, therefore, in the space of *rational* spectral densities  $\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$ . Interestingly, it turns out that in the latter case the expressions of Thompson and Hilbert metric are connected with a classical problem in systems theory, the spectral factorization problem.

*Theorem 1:* Consider two full normal rank spectral densities  $\Phi_1, \Phi_2 \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  and let  $W_1, W_2 \in \mathbb{R}^{n \times n}(z)$  denote the corresponding minimum-phase spectral factors. If  $W_2^{-1}W_1$  has no zero/pole on  $\mathbb{T}$ , then the Hilbert and Thompson metrics between  $\Phi_1$  and  $\Phi_2$  are given, respectively, by

$$d_H(\Phi_1, \Phi_2) = \log \|W_2^{-1}W_1\|_{\mathcal{H}_\infty}^2 \|W_1^{-1}W_2\|_{\mathcal{H}_\infty}^2, \\ d_T(\Phi_1, \Phi_2) = \log \max \left\{ \|W_2^{-1}W_1\|_{\mathcal{H}_\infty}^2, \|W_1^{-1}W_2\|_{\mathcal{H}_\infty}^2 \right\}.$$

Otherwise, it holds  $d_H(\Phi_1, \Phi_2) = d_T(\Phi_1, \Phi_2) = \infty$ .

*Proof:* In view of the definition of  $M(\cdot, \cdot)$  in (2), for any full normal rank  $\Phi_1, \Phi_2 \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$ , it holds

$$M(\Phi_1, \Phi_2) = \inf \{ \lambda \in \mathbb{R} : \Phi_1(e^{j\theta}) \leq \lambda \Phi_2(e^{j\theta}), \\ \theta \in [-\pi, \pi] \} \\ = \inf \{ \lambda \in \mathbb{R} : \Phi_2^{-\frac{1}{2}}(e^{j\theta}) \Phi_1(e^{j\theta}) \Phi_2^{-\frac{1}{2}}(e^{j\theta}) \}$$

$$\leq \lambda I_n, \theta \in [-\pi, \pi] \}$$

$$= \left\| \Phi_2^{-\frac{1}{2}} \Phi_1 \Phi_2^{-\frac{1}{2}} \right\|_{\mathcal{L}_\infty}, \quad (10)$$

if  $\Phi_2^{-\frac{1}{2}} \Phi_1 \Phi_2^{-\frac{1}{2}}$  is analytic on  $\mathbb{T}$ , and  $M(\Phi_1, \Phi_2) = \infty$  otherwise. In order to deal with rational matrix-valued functions we can replace, without affecting the value of  $M(\Phi_1, \Phi_2)$ , the square root  $\Phi_2^{1/2}$  in the latter expression with the minimum-phase spectral factor  $W_2 \in \mathbb{R}^{n \times n}(z)$  of  $\Phi_2$ . (In fact,  $\Phi_2^{1/2} U = W_2$ , where  $U$  is a suitable  $n \times n$  unitary matrix-valued function on  $\mathbb{T}$ ). Therefore, Equation (10) becomes

$$M(\Phi_1, \Phi_2) = \left\| \Phi_2^{-\frac{1}{2}} \Phi_1 \Phi_2^{-\frac{1}{2}} \right\|_{\mathcal{L}_\infty} \\ = \left\| W_2^{-1} \Phi_1 W_2^{-*} \right\|_{\mathcal{L}_\infty} \\ = \left\| W_2^{-1} W_1 W_1^* W_2^{-*} \right\|_{\mathcal{L}_\infty} \\ = \left\| W_2^{-1} W_1 \right\|_{\mathcal{L}_\infty}^2,$$

if  $W_2^{-1}W_1$  has no pole on  $\mathbb{T}$ , and  $M(\Phi_1, \Phi_2) = \infty$  otherwise. Further, if  $W_2^{-1}W_1$  has no pole on  $\mathbb{T}$ ,  $W_2^{-1}W_1$  is analytic in (an open set containing) the complement of  $\mathbb{D}$ , so that we have

$$M(\Phi_1, \Phi_2) = \left\| W_2^{-1}W_1 \right\|_{\mathcal{H}_\infty}^2,$$

where we have replaced the  $\mathcal{L}_\infty$ -norm with the  $\mathcal{H}_\infty$ -norm. Similarly, we have

$$M(\Phi_2, \Phi_1) = \left\| W_1^{-1}W_2 \right\|_{\mathcal{H}_\infty}^2,$$

if  $W_1^{-1}W_2$  has no pole on  $\mathbb{T}$ , or, equivalently, if  $W_2^{-1}W_1$  has no zero on  $\mathbb{T}$ , and  $M(\Phi_2, \Phi_1) = \infty$  otherwise.

Eventually, observing that  $m(\Phi_1, \Phi_2) = M(\Phi_2, \Phi_1)^{-1}$ , a substitution of the values of  $M(\Phi_1, \Phi_2)$  and  $M(\Phi_2, \Phi_1)$  into the expressions of Hilbert and Thompson metrics in Definition 1 yields the thesis. ■

*Remark 1:* The proof of Theorem 1 shows that the expressions of the Hilbert and Thompson metric still hold if we replace the canonical (i.e., minimum-phase) spectral factors of the two spectra  $\Phi_1, \Phi_2$  with any other spectral factor of  $\Phi_1, \Phi_2$  (i.e., spectral factors not necessarily analytic in the complement of  $\mathbb{D}$  and with analytic inverse in the complement of the closure of  $\mathbb{D}$ ). The important difference is that, in this case, the  $\mathcal{H}_\infty$ -norm must be replaced by the  $\mathcal{L}_\infty$ -norm.

*Remark 2:* As discussed in the previous section, the difference between Hilbert and Thompson metric consists of the fact that the Thompson metric is a bona fide distance on each part of  $\mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  (and, in particular, on its interior), while Hilbert metric is a distance between rays in each part of the latter cone. It is worth remarking that projective invariance has proved to be a desirable property since in many applications, such as spectral estimation or speech processing, the shape of the spectral densities rather than their relative scalings is the discriminative feature [2], [38].

*Remark 3:* The expressions of Hilbert and Thompson metrics in Theorem 1 apply also to the case of general non-rational spectral densities in  $\mathcal{S}_+^{n \times n}(\mathbb{T})$ .<sup>3</sup> In this case, however, one issue

<sup>3</sup>Notice that in case the minimum-phase spectral factors of  $\Phi_1, \Phi_2$  do not exist, the expressions in Theorem 1 still holds by replacing the minimum-phase spectral factors with the corresponding frequency-wise matrix square roots  $\Phi_1^{1/2}, \Phi_2^{1/2}$ .

that arises is that the distance between almost identical spectral densities can be made arbitrarily large. With reference to the scalar case, this occurs when one of the two spectral densities exhibits a very sharp and narrow frequency peak. For the sake of illustration, consider the two scalar spectral densities in  $\mathcal{S}_{+}^{1 \times 1}(\mathbb{T})$

$$\phi_1(e^{j\theta}) = 1, \quad \phi_{2,\varepsilon}(e^{j\theta}) = \begin{cases} \varepsilon^{-1} & \text{if } |\theta| \leq \varepsilon, \\ 1 & \text{otherwise,} \end{cases} \quad (11)$$

where  $\theta \in [-\pi, \pi]$  and  $\varepsilon > 0$ . It can be seen that, for  $\varepsilon \rightarrow 0$ ,  $d_H(\phi_1, \phi_{2,\varepsilon}) \rightarrow \infty$  and  $d_T(\phi_1, \phi_{2,\varepsilon}) \rightarrow \infty$ , in spite of the fact that the two spectral densities are identical with the only exception of a neighborhood of the frequency  $\theta = 0$  (see also Figure 1). Importantly, when restricting the attention to spectral densities that are “sufficiently regular”, e.g., those belonging to the space of rational spectral densities with bounded McMillan degree, these pathological cases are ruled out.

In view of the above property, the proposed distances are not suitable to treat spectral densities featuring “spectral lines”. On the other hand, the same property is a relevant feature in a distance for the THREE spectral estimation problem described in Example 3. Indeed, in view of this property the THREE problem with a distance as the ones introduced in this paper cannot produce solutions featuring “artifacts” (spurious spectral lines) which is one of the main drawbacks of the method with classical distances. In this sense the distances just introduced can be used for a “robust” version of the THREE problem (see also the concluding remarks for more details).

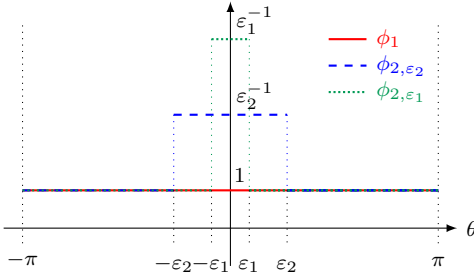


Fig. 1. Qualitative plot of the spectral densities in (11) for two values of  $\varepsilon$ , i.e.  $\varepsilon_2 > \varepsilon_1 > 0$ .

#### A. Filtering invariance

The above-introduced distances possess the following important property:

$$\forall T \in \mathbb{R}_*^{n \times n}(z) : d(\Phi_1, \Phi_2) = d(T\Phi_1 T^*, T\Phi_2 T^*). \quad (12)$$

This property readily follows from the definition of Hilbert and Thompson distances and the expression of  $M(\Phi_1, \Phi_2)$  in (10). Since the set  $\mathbb{R}_*^{n \times n}(z)$  defines a group, the mapping  $\Phi \mapsto T\Phi T^*$  defines a congruence group action of  $\mathbb{R}_*^{n \times n}(z)$  on the set of rational spectral densities. This group action is transitive, that is, any rational spectral density can be obtained by acting on the identity element.

A metric that satisfies (12) can be said to be *filtering invariant* because of the following statistical interpretation.

Any spectral density  $\Phi$  with minimum-phase spectral factor  $W$  can be identified to a  $n$ -dimensional zero-mean second-order stationary purely nondeterministic stochastic process  $\{y(t)\}_{t \in \mathbb{Z}}$  generated by filtering a white noise process through  $W$ . The action  $\Phi \mapsto T\Phi T^*$  has therefore the interpretation of filtering the process with the linear time-invariant filter  $T \in \mathbb{R}_*^{n \times n}(z)$ . Likewise, the property (12) has the interpretation that the distance between two spectral densities, or, equivalently, two zero-mean second-order stationary purely nondeterministic stochastic processes, is unchanged when the two processes are filtered by the same filter.

Any filtering invariant metric is entirely specified by defining the distance to identity. Furthermore, one has  $d(\Phi, I) = d(\Phi^{-1}, I)$ . In other words, the distance is a *distortion* measure.

Filtering invariance is a fundamental property of classical metrics. In the scalar case,  $\phi_1, \phi_2 \in \mathcal{S}_{+, \text{rat}}^{1 \times 1}(\mathbb{T})$ , the log spectral deviation [11]

$$\left( \int_{-\pi}^{\pi} \left\| \log \frac{\phi_1}{\phi_2} \right\|^2 \frac{d\theta}{2\pi} \right)^{1/2}$$

is an early example of filtering invariant distortion measure. The recent work [23] shows that the multivariate generalization

$$\begin{aligned} d_R(\Phi_1, \Phi_2) &= \left( \int_{-\pi}^{\pi} \left\| \log \Phi_1^{-1/2} \Phi_2 \Phi_1^{-1/2} \right\|_F^2 \frac{d\theta}{2\pi} \right)^{1/2} \\ &= \left( \int_{-\pi}^{\pi} \left\| \log W_1^{-1} \Phi_2 W_1^{-*} \right\|_F^2 \frac{d\theta}{2\pi} \right)^{1/2} \end{aligned} \quad (13)$$

is the unique Riemannian bona fide distance that is filtering invariant. This metric is a natural generalization of the affine invariant metric between positive definite matrices. Affine invariance corresponds to filtering invariance in the static case: the congruence group action reduces to an action of the general linear group. The metric is in this case a distance between  $n$ -dimensional zero-mean second-order random vectors, and the invariance property is an invariance with respect to an affine change of coordinates. The importance of this invariance property in the context of estimation problems has been emphasized for instance in [39]. In [23], filtering invariance emerges as a natural property when measuring the “flatness” of innovations processes. Filtering invariance is also a leading prerequisite in the work of Martin [40], whose resulting metric, which applies to scalar spectral densities  $\phi_1, \phi_2 \in \mathcal{S}_{+, \text{rat}}^{1 \times 1}(\mathbb{T})$ , can be written as

$$d_M(\phi_1, \phi_2) = \left( \int_{-\pi}^{\pi} \left( \mathfrak{D}^{\frac{1}{2}} \log \frac{\phi_1}{\phi_2} \right)^2 \frac{d\theta}{2\pi} \right)^{1/2}, \quad (14)$$

where  $\mathfrak{D}^\lambda$ ,  $\lambda > 0$ , is the fractional derivative operator in the frequency domain. Finally, filtering invariance is also a key property of the classical Itakura–Saito divergence (see [5], [41]).

The Riemannian distance (13) and the Thompson metric introduced in Section 2 are thus close relatives: they are bona fide distances which satisfy filtering invariance and endow the cone of spectral densities with a differential metric structure. The first one induces a Riemannian structure through an invariant inner product (that reduces to the standard inner

product at identity), while the second induces a Finslerian structure through the invariant norm (6). Both distances depend on the same log spectral quantity frequency-wise, but the Riemannian distance results in a two-norm of that frequency-domain function, whereas the Finsler distance results in an infinite-norm.

### B. Computational properties

Theorem 1 shows that the computation of Hilbert and Thompson metrics in the cone of rational spectral densities essentially consists of: (i) the calculation of the minimum-phase spectral factors  $W_1$  and  $W_2$ , and (ii) the calculation of the  $\mathcal{H}_\infty$ -norm of the “ratio” of the latter spectral factors. For the sake of clarity, we illustrate below a simple example of computation of these metrics.

*Example 4:* Consider the scalar rational spectral densities

$$\phi_1(z) = -\frac{4}{2z + -5 + 2z^{-1}},$$

$$\phi_2(z) = \frac{9}{3z + 10 + 3z^{-1}}.$$

The minimum-phase spectral factors of  $\phi_1$  and  $\phi_2$  are given by, respectively,

$$w_1(z) = \frac{z}{z - 1/2}, \quad w_2(z) = \frac{z}{z + 1/3}.$$

By direct computation, we have

$$\left\| \frac{w_2}{w_1} \right\|_{\mathcal{H}_\infty}^2 = \sup_{|z|>1} \frac{|z + 1/3|}{|z - 1/2|} = \frac{8}{3}$$

$$\left\| \frac{w_1}{w_2} \right\|_{\mathcal{H}_\infty}^2 = \sup_{|z|>1} \frac{|z - 1/2|}{|z + 1/3|} = \frac{9}{4}.$$

Hence, in view of Theorem 1,

$$d_H(\phi_1, \phi_2) = \log \left\| \frac{w_2}{w_1} \right\|_{\mathcal{H}_\infty}^2 \left\| \frac{w_1}{w_2} \right\|_{\mathcal{H}_\infty}^2 = \log 6,$$

$$d_T(\phi_1, \phi_2) = \log \max \left\{ \left\| \frac{w_2}{w_1} \right\|_{\mathcal{H}_\infty}^2, \left\| \frac{w_1}{w_2} \right\|_{\mathcal{H}_\infty}^2 \right\} = \log \frac{8}{3}.$$

Remarkably, the computation of (minimum-phase) spectral factors and  $\mathcal{H}_\infty$ -norms represent two extensively studied problems in systems and control theory and several algorithms are available in the literature to perform these steps, even in the most challenging multivariate case. More specifically, a general result on the existence and (essential) uniqueness of minimum-phase spectral factors can be found in [42] for the continuous-time case and in [25] for the discrete-time case. An algorithm to compute this spectral factor is described there, however, for the calculation of minimum-phase spectral factors there exist several more efficient routines based on the solution of suitable Stein and Riccati Equations, see e.g. [43]. Whereas, an efficient method for computing  $\mathcal{H}_\infty/\mathcal{L}_\infty$ -norms of rational matrix-valued transfer functions is the Boyd–Balakrishnan–Bruinsma–Steinbuch method [44], [45] which is based on an iterative bisection-like algorithm and leads to quadratic

convergence.<sup>4</sup> Moreover, an upper bound to these norms can be found by inspecting the eigenvalues of the symplectic matrix associated with the state-space representation of the system [47, Lemma 21.10].

In view of the previous analysis, a merit of the Finslerian (Thompson) distance over its Riemannian relative (13) is at the computational level. In fact, the calculation of the Riemannian distance requires the frequency-wise computation of the (matrix) logarithm of  $W_1^{-1}\Phi_2W_1^{-*}$ , an operation which appears numerically challenging in the multivariate setting. In contrast, the calculation of the Thompson metric involves the computation of minimum-phase spectral factors and  $\mathcal{H}_\infty$ -norms, for which efficient numerical algorithms are available.

For completeness, it should be mentioned that one way of overcoming the computational burden of the Riemannian distance is to replace it with a divergence measure. In the static case, Kullback–Leibler divergence approximates the Riemannian distance up to third order. In the dynamic case, the paper [23] considers quadratic approximations of divergence measures. In the rational case, one such quantity takes the form

$$d_F(\Phi_1, \Phi_2) = \|W_2^{-1}W_1\|_{\mathcal{H}_2}^2 + \|W_1^{-1}W_2\|_{\mathcal{H}_2}^2 - 2n,$$

where  $\|\cdot\|_{\mathcal{H}_2}$  denotes the  $\mathcal{H}_2$ -norm of a discrete-time transfer function [47, Sec. 4.3]. It is not a distance (in fact, it does not obey the triangle inequality) but it provides a tractable quadratic approximation of the Riemannian distance.

### C. Geodesic paths

One can generalize the geodesic expression (8) obtained for the positive definite matrix case to the case of bounded positive operators on an Hilbert space [34], [35]. From this generalization, it follows that a minimal geodesic path in  $\mathcal{S}_+^{n \times n}(\mathbb{T})$  connecting  $\Phi_1, \Phi_2 \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  w.r.t. Thompson metric is given by<sup>5</sup>

$$\varphi_T(\tau) = W_1(W_1^{-1}\Phi_2W_1^{-*})^\tau W_1^*, \quad \tau \in [0, 1], \quad (15)$$

where  $W_1 \in \mathbb{R}^{n \times n}(z)$  is the minimum-phase spectral factor of  $\Phi_1$ . Notice that Equation (15) coincides with the (unique, up to a re-parametrization) Riemannian geodesic between spectral densities [23].

However, as discussed in Sec. III, the Finslerian framework allows for the definition of multiple minimal geodesic paths. For instance, an alternative explicit Thompson’s minimal geodesic for spectral densities follows the “projective” straight line interpolation in (7). This yields the following minimal geodesic path connecting  $\Phi_1, \Phi_2 \in \mathcal{S}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  w.r.t. Thompson metric

$$\chi_T(\tau) = \begin{cases} \left( \frac{\beta^\tau - \alpha^\tau}{\beta - \alpha} \right) \Phi_2 + \left( \frac{\beta\alpha^\tau - \alpha\beta^\tau}{\beta - \alpha} \right) \Phi_1, & \text{if } \beta \neq \alpha, \\ \alpha^\tau \Phi_1, & \text{if } \beta = \alpha, \end{cases} \quad (16)$$

<sup>4</sup>The approach is formulated in the continuous-time case, however there exist variants of this algorithm for computing the discrete-time  $\mathcal{H}_\infty$ -norm, based on computing eigenvalues of symplectic instead of Hamiltonian matrices, see e.g. [46].

<sup>5</sup>Notice that we have expressed the geodesic in terms of the rational spectral factor  $W_1$  instead of the frequency-wise matrix square root  $\Phi_1^{1/2}$ . This equivalent rewriting follows from the fact that  $\Phi_1^{1/2} = W_1U$  with  $U$  being an  $n \times n$  unitary matrix-valued function on  $\mathbb{T}$ .

where  $\tau \in [0, 1]$ ,

$$\beta := \|W_1^{-1}W_2\|_{\mathcal{H}_\infty}^2, \quad \alpha := 1/\|W_2^{-1}W_1\|_{\mathcal{H}_\infty}^2,$$

with  $W_1 \in \mathbb{R}^{n \times n}(z)$  and  $W_2 \in \mathbb{R}^{n \times n}(z)$  being the minimum-phase spectral factors of  $\Phi_1$  and  $\Phi_2$ , respectively.

When applied to the whole cone  $\mathring{\mathcal{S}}_+^{n \times n}(\mathbb{T})$ , the Riemannian geodesic path (15) renders this space geodesically complete, meaning that for all  $\tau \in \mathbb{R}$ ,  $\varphi_T(\tau)$  belongs to  $\mathring{\mathcal{S}}_+^{n \times n}(\mathbb{T})$ . This is indeed a remarkable property that allows for extrapolating along the geodesic paths. On the other hand, when considering rational spectral densities,  $\varphi_T(\tau)$  is not, in general, rational. Concerning the Finslerian geodesic path in (16) we have the following result.

**Proposition 1:** For all  $\Phi_1, \Phi_2 \in \mathring{\mathcal{S}}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  and  $\tau \in \mathbb{R}$ ,  $\chi_T(\tau) \in \mathring{\mathcal{S}}_{+, \text{rat}}^{n \times n}(\mathbb{T})$ .

*Proof:* The path  $\chi_T(\tau)$  is bounded and rational for all  $\tau \in \mathbb{R}$  by construction since  $\alpha$  and  $\beta$  are bounded real scalars. So it remains to prove positivity of  $\chi_T(\tau)$ . The case  $\beta = \alpha$  is straightforward, so in what follows we suppose  $\beta \neq \alpha$ . Notice that, by pre- and post-multiplying  $\chi_T(\tau)$  by  $\Phi^{-1/2}$  and then diagonalizing the resulting expression, the condition  $\chi_T(\tau) > 0$  for all  $\theta \in [-\pi, \pi]$  can be seen to be equivalent to

$$\left( \frac{\beta^\tau - \alpha^\tau}{\beta - \alpha} \right) \Lambda(e^{j\theta}) + \left( \frac{\beta\alpha^\tau - \alpha\beta^\tau}{\beta - \alpha} \right) I > 0, \quad (17)$$

for all  $\theta \in [-\pi, \pi]$ , where

$$\Lambda(e^{j\theta}) := \text{diag}[\lambda_1(e^{j\theta}), \lambda_2(e^{j\theta}), \dots, \lambda_n(e^{j\theta})],$$

$\lambda_1(e^{j\theta}) \geq \lambda_2(e^{j\theta}) \geq \dots \geq \lambda_n(e^{j\theta}) > 0$ ,  $\theta \in [-\pi, \pi]$ , has in its diagonal the frequency-wise eigenvalues of  $\Phi_1^{-1/2}(e^{j\theta})\Phi_2(e^{j\theta})\Phi_1^{-1/2}(e^{j\theta})$ . Next, we note that

$$\lambda_n(e^{j\theta}) \geq \min_{\theta \in [-\pi, \pi]} \lambda_n(e^{j\theta}) = 1/\|W_2^{-1}W_1\|_{\mathcal{H}_\infty}^2 = \alpha.$$

In view of Eq. (17), this in turn implies that for all  $\theta \in [-\pi, \pi]$ ,

$$\begin{aligned} & \left( \frac{\beta^\tau - \alpha^\tau}{\beta - \alpha} \right) \Lambda(e^{j\theta}) + \left( \frac{\beta\alpha^\tau - \alpha\beta^\tau}{\beta - \alpha} \right) I \\ & \geq \left( \frac{\beta^\tau - \alpha^\tau}{\beta - \alpha} \right) \alpha I + \left( \frac{\beta\alpha^\tau - \alpha\beta^\tau}{\beta - \alpha} \right) I = \alpha^\tau I > 0, \end{aligned}$$

which completes the proof.  $\blacksquare$

In more practical terms, the previous proposition states that one can interpolate along the geodesic path (16) while remaining in the cone of *rational* spectral densities. This is a fundamental feature of Finsler geometry that does not have a Riemannian counterpart. In light of this fact, we argue that, when dealing with rational spectral densities, the Finsler geodesic (16) may be a more natural choice when compared to the Riemannian geodesic (15). In view of the geodesical completeness, this is true also for *extrapolation* so that, given two rational spectral densities  $\Phi_1$  and  $\Phi_2$ , we can select a rational spectral density in the geodesic line connecting  $\Phi_1$  and  $\Phi_2$  and this spectral density is not necessarily between  $\Phi_1$  and  $\Phi_2$  but may also be chosen to be “before”  $\Phi_1$  or “after”  $\Phi_2$ . This is particularly interesting in applications as illustrated in the next section where the morphing between a male and a female voice is discussed and we can, for example, go “beyond male” and synthesize a particularly baritone voice.

**Remark 4:** Geodesic expressions (15) and (16) for Thompson metric can be adapted to geodesics for Hilbert metric by considering the corresponding “normalized” versions. For instance, w.r.t. “normalized” spectra  $\Phi_1, \Phi_2 \in \mathring{\mathcal{S}}_{+, \text{rat}}^{n \times n}(\mathbb{T})$  such that  $\int_{-\pi}^{\pi} \text{tr}(\Phi_1) \frac{d\theta}{2\pi} = \int_{-\pi}^{\pi} \text{tr}(\Phi_2) \frac{d\theta}{2\pi} = 1$ , geodesic (15) becomes

$$\varphi_H(\tau) = \frac{W_1(W_1^{-1}\Phi_2W_1^{-*})^\tau W_1^*}{\int_{-\pi}^{\pi} \text{tr}(W_1(W_1^{-1}\Phi_2W_1^{-*})^\tau W_1^*) \frac{d\theta}{2\pi}}, \quad \tau \in [0, 1]. \quad (18)$$

**Remark 5:** The reader will observe that the geodesic paths discussed in this section inherit the invariance properties of the metric discussed in Subsection IV-A. Hence, the construction of geodesic connecting curves between any two points can always be recast as the construction of a geodesic connecting curve between an arbitrary point and the identity.

## V. APPLICATIONS: SPEECH MORPHING

In this section, we show how to apply the geometric tools we developed in the previous section as a means to interpolate and extrapolate rational spectra describing the frequency content of speech data. More precisely, we analyze the task of morphing the voice of an individual into the voice of another individual. To this end, we first review some standard facts concerning speech modelling and synthesis, that can be found, for instance, in [48]–[50].

Speech signals can be considered approximately stationary when restricted to a small time interval (typically,  $\sim 25$  ms). Within such an interval, a speech signal can be modelled by a linear time-invariant filter driven by a suitable excitation signal. For each time fragment, the morphing of two speech signals can be accomplished via interpolation of the (rational) spectral densities describing the two modelled signals. Starting from the morphed spectrum, a morphed speech signal can be generated as an output of suitable linear filter, similarly as before. The complete morphed speech signal can be eventually recovered by stacking together all the morphed audio fragments.

As suggested in [13] and briefly mentioned in Section II, the task of interpolating two spectral densities can be naturally and efficiently carried out using suitable geodesic paths. Here, we focus on spectral interpolation via the Finslerian geodesic in (16).

Before illustrating the obtained results, we briefly discuss the implementation details of the morphing procedure and how the latter compares to the method proposed in [13].

The morphing approach we considered is schematically depicted in the block diagram of Figure 2. Here, we consider two audio samples  $s_M(t)$  and  $s_W(t)$  corresponding to the phoneme /a:/ spoken by a male and a female individual, respectively. Each audio signal is sampled at 16 kHz and has a duration of 0.3 s. We partition the signals into frames of 25 ms, and we estimate the pitch period of the male speech signal ( $p_M$ ) and of the female speech signal ( $p_W$ ) via residual-based estimation [50, Sec. 6.7]. As common practice, signals  $s_M(t)$  and  $s_W(t)$  are pre-filtered with a “pre-emphasis” filter in order to reduce their low-frequency



content. Then an Hamming window convolution is applied to each filtered signal. The linear model estimation (known also as linear predictive coding) is performed using classical techniques: first we estimate the covariance lags of the signals via auto-correlation method and then we apply the Levinson–Durbin method to obtain the AR model coefficients<sup>6</sup> [50, Sec. 6.3]. From the linear model the corresponding rational power spectral densities of the estimated signals, denoted by  $\phi_M(e^{j\theta})$  and  $\phi_W(e^{j\theta})$ , are computed and then interpolated using Finsler geodesic (16). The resulting rational morphed spectrum is denoted by  $\phi_\tau(e^{j\theta})$ ,  $\tau \in [0, 1]$ , where  $\phi_0 := \phi_W$  and  $\phi_1 := \phi_M$ . The synthesis of the morphed speech signal is then simply performed by feeding the minimum-phase rational spectral factor of the morphed spectrum  $\phi_\tau(e^{j\theta})$  with an excitation signal consisting of a pulse train with frequency  $p_\tau$ , where  $p_\tau$  is obtained by linear or geometric interpolation of  $p_M$  and  $p_W$ . Finally, a “post-emphasis” filter is applied to the resulting speech signal in order to compensate the effect of the “pre-emphasis” filter.

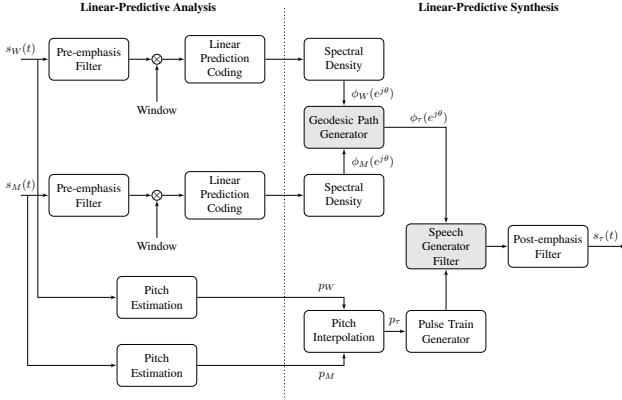


Fig. 2. Block diagram of the speech morphing procedure. The scheme is divided in two main blocks: Linear-Predictive Analysis wherein a linear AR model for each speech signal is estimated, and Linear-Predictive Synthesis wherein the morphed speech is synthesized. The main differences w.r.t. the morphing method proposed in [13] are highlighted in gray.

The above illustrated morphing scheme shares many similarities with the procedure adopted in [13]. However, a main difference is apparent. In [13] the synthesis step requires the estimation of a (AR) linear model from the (typically non-rational) morphed spectrum. Using Finslerian geodesic (16) this additional step can be bypassed since  $\phi_\tau(e^{j\theta})$  is rational by construction, allowing for a (considerable) simplification of the digital implementation procedure.

The top plot of Figure 3 shows the temporal behavior of the two considered speech waveforms  $s_M(t)$  and  $s_W(t)$ . The dashed gray zone in the top plot highlights a specific time fragment. With reference to this time fragment, the bottom plot of Figure 3 shows the two rational spectral densities estimated using the linear predictive estimation procedure outlined above.

In Figure 4, the result of interpolation between  $\phi_M(e^{j\theta})$  and  $\phi_W(e^{j\theta})$  via Finslerian geodesic (16) is depicted. The dashed

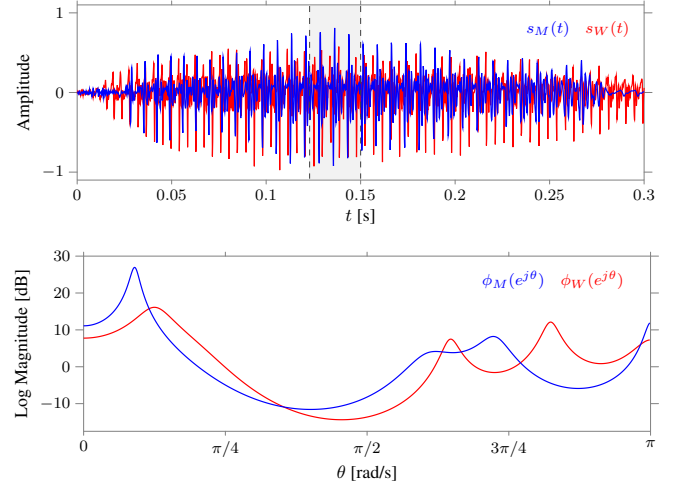


Fig. 3. Top plot: Speech signals corresponding to phoneme /a:/ spoken by a male ( $s_M(t)$ ) and female individual ( $s_W(t)$ ). Bottom plot: (One-side) rational spectral densities obtained from linear predictive modelling of  $s_M(t)$  and  $s_W(t)$  within the time interval highlighted in the dashed gray zone in the top plot.

curves denote spectral densities obtained by extrapolating along the geodesic path. From this figure, it is interesting to observe that the interpolation/extrapolation behavior of the peaks of the morphed spectral density seems to be almost linear in a logarithmic scale.

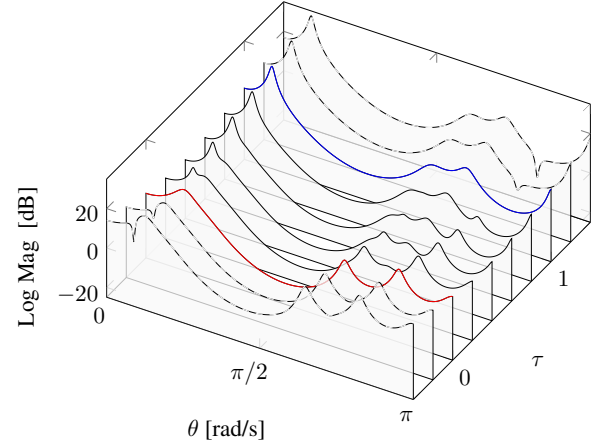


Fig. 4. Interpolation of  $\phi_M(e^{j\theta})$  (blue curve) and  $\phi_W(e^{j\theta})$  (red curve) via the Finslerian geodesic (16) for different values of  $\tau \in \mathbb{R}$ . The dashed curves denote extrapolated spectra.

Results of morphing for full words and sentences are available in an audible format in [51]. In spite of the quite rudimentary nature of the procedure and the fact that the effect of the acoustic quality is very subjective, the obtained morphed audio signal seems often to be of acceptable quality. Of course, the proposed morphed procedure can be thought of as a building block which allows for the incorporation of more sophisticated processing tools to enhance the quality of the resulting speech signal.

<sup>6</sup>Here, we fixed the AR model order to 14.



## VI. CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this paper, we studied a class of conal distances for rational spectral densities arising from Finsler geometry. The proposed Thompson and Hilbert metric have a number of attractive and unique features. In particular, they are easily computable, they enjoy filtering invariance and they possess minimal geodesics that preserve rationality. These properties make these distances suitable for application in a variety of problems across systems and control theory and signal processing.

For instance, a problem that could benefit from the use of such distances is spectral estimation in the THREE-like framework discussed in the introductory Example 3. A common feature of all distances proposed so far to tackle this problem is that they involve the two-norm of a frequency-wise quantity defined on the unit circle. Choosing the Finsler distances of this paper could lead to a *robust* version of THREE-like spectral estimation. A main motivation for this modified formulation concerns the reduction of *artifacts* in the solution of this problem. The presence of artifacts is an issue that affects many of the spectral estimation methods proposed in the literature (see, e.g., [5, Sec. VII-B]). Artifacts are usually present in the form of high and narrow frequency peaks in the spectral estimate. In light of Remark 3, the advantage of using either Thompson or Hilbert distance in solving the spectral estimation problem consists of the fact that these artifacts are highly penalized by these distances, and, consequently, they should not appear in the optimal spectral estimate.

More generally, the optimization of  $\mathcal{H}_\infty$ -norms in place of or in complement to  $\mathcal{H}_2$ -norms has been a very fruitful direction of research in linear system theory. Building upon this heritage, the distances introduced in the present paper could also open novel avenues in robust statistical estimation. This program raises a number of open questions, as Finsler optimization is a far less mature research area than Riemannian optimization.

Finally, Finsler distances are also a promising tool to introduce a geometry in the set of passive systems whose transfer functions are *positive real* and hence are in one-to-one correspondence with a spectral density. Negative imaginary systems may inherit this geometry in view of the connections between positive real and negative imaginary transfer functions [52], [53].

## REFERENCES

- [1] M. Basseville, "Divergence measures for statistical data processing – an annotated bibliography," *Signal Processing*, vol. 93, no. 4, pp. 621–633, 2013.
- [2] B. Afsari and R. Vidal, "Distances on spaces of high-dimensional linear stochastic processes: A survey," in *Geometric Theory of Information*. Springer, 2014, pp. 219–242.
- [3] T. T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 2910–2917, 2003.
- [4] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger versus Kullback–Leibler multivariable spectrum approximation," *IEEE Transactions on Automatic Control*, vol. 53, no. 4, pp. 954–967, 2008.
- [5] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2561–2575, 2012.
- [6] P. Enqvist and J. Karlsson, "Minimal Itakura-Saito distance and covariance interpolation," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*. IEEE, 2008, pp. 137–142.
- [7] P. Enqvist, "Covariance interpolation and geometry of power spectral densities," in *Control Conference (ECC), 2009 European*. IEEE, 2009, pp. 4505–4510.
- [8] M. Zorzi, "A new family of high-resolution multivariate spectral estimators," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 892–904, April 2014.
- [9] —, "Rational approximations of spectral densities based on the alpha divergence," *Mathematics of Control, Signals, and Systems*, vol. 26, no. 2, pp. 259–278, 2014.
- [10] —, "Multivariate spectral estimation based on the concept of optimal prediction," *IEEE Transactions on Automatic Control*, vol. 60, no. 6, pp. 1647–1652, June 2015.
- [11] R. Gray, A. Buzo, A. Gray, and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 367–376, 1980.
- [12] V. Goncharoff and M. Kaine-Krolak, "Interpolation of LPC spectra via pole shifting," in *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 1. IEEE, 1995, pp. 780–783.
- [13] X. Jiang, S. Takyar, and T. T. Georgiou, "Metrics and morphing of power spectra," in *Recent Advances in Learning and Control*. Springer, 2008, pp. 125–135.
- [14] X. Jiang, Z.-Q. Luo, and T. T. Georgiou, "Power spectral geodesics and tracking," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*. IEEE, 2008, pp. 1315–1319.
- [15] X. Jiang, J. Karlsson, and T. T. Georgiou, "Phoneme segmentation based on spectral metrics," in *International Symposium on Mathematical Theory of Networks and Systems*, 2008.
- [16] X. Jiang, Z.-Q. Luo, and T. T. Georgiou, "Geometric methods for spectral analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1064–1074, 2012.
- [17] T. W. Liao, "Clustering of time series data — a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
- [18] J. Boets, K. De Cock, B. De Moor, and M. Espinoza, "Clustering time series, subspace identification and cepstral distances," *Communications in Information & Systems*, vol. 5, no. 1, pp. 69–96, 2005.
- [19] J. Boets, K. De Cock, and B. De Moor, "Distances between dynamical models for clustering time series," *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 392–397, 2006.
- [20] M. Corduas and D. Piccolo, "Time series clustering and classification by the autoregressive metric," *Computational statistics & data analysis*, vol. 52, no. 4, pp. 1860–1872, 2008.
- [21] B. Afsari, R. Chaudhry, A. Ravichandran, and R. Vidal, "Group action induced distances for averaging and clustering linear dynamical systems with applications to the analysis of dynamic scenes," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2208–2215.
- [22] O. Lauwers and B. De Moor, "A time series distance measure for efficient clustering of input/output signals by their underlying dynamics," *IEEE Control Systems Letters*, vol. 1, no. 2, pp. 286–291, 2017.
- [23] X. Jiang, L. Ning, and T. T. Georgiou, "Distances and Riemannian metrics for multivariate spectral densities," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1723–1735, 2012.
- [24] A. Lindquist and G. Picci, *Linear stochastic systems: A Geometric Approach to Modeling, Estimation and Identification*. Springer, 2015.
- [25] G. Baggio and A. Ferrante, "On the factorization of rational discrete-time spectral densities," *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 969–981, 2016.
- [26] P. J. Bushell, "Hilbert's metric and positive contraction mappings in a Banach space," *Archive for Rational Mechanics and Analysis*, vol. 52, no. 4, pp. 330–338, 1973.
- [27] A. Thompson, "On certain contraction mappings in a partially ordered vector space," *Proceedings of the American Mathematical Society*, vol. 14, no. 3, pp. 438–443, 1963.
- [28] C. Liverani and M. P. Wojtkowski, "Generalization of the Hilbert metric to the space of positive definite matrices," *Pacific J. Math*, vol. 166, no. 2, pp. 339–355, 1994.
- [29] B. Lemmens and R. Nussbaum, *Nonlinear Perron-Frobenius Theory*. Cambridge University Press, 2012, vol. 189.
- [30] E. Kohlberg and J. W. Pratt, "The contraction mapping approach to the Perron-Frobenius theory: Why Hilbert's metric?" *Mathematics of Operations Research*, vol. 7, no. 2, pp. 198–210, 1982.

- [31] R. D. Nussbaum, "Finsler structures for the part metric and Hilbert's projective metric and applications to ordinary differential equations," *Differential and Integral Equations*, vol. 7, no. 5-6, pp. 1649-1707, 1994.
- [32] R. D. Nussbaum and C. Walsh, "A metric inequality for the Thompson and Hilbert geometries," *J. Inequal. Pure Appl. Math*, vol. 5, no. 3, 2004.
- [33] R. Bhatia, *Positive definite matrices*. Princeton university press, 2009.
- [34] G. Corach, H. Porta, and L. Recht, "Geodesics and operator means in the space of positive operators," *International Journal of Mathematics*, vol. 4, no. 02, pp. 193-202, 1993.
- [35] —, "Convexity of the geodesic distance on spaces of positive operators," *Illinois J. Math*, vol. 38, no. 1, pp. 87-94, 1994.
- [36] G. Corach and A. L. Maestripietri, "Differential and metrical structure of positive operators," *Positivity*, vol. 3, no. 4, pp. 297-315, 1999.
- [37] —, "Differential geometry on Thompson's components of positive operators," *Reports on Mathematical Physics*, vol. 45, no. 1, pp. 23-37, 2000.
- [38] T. T. Georgiou, "Distances and Riemannian metrics for spectral density functions," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 3995-4003, 2007.
- [39] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610-1630, 2005.
- [40] R. J. Martin, "A metric for ARMA processes," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1164-1170, 2000.
- [41] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electronics and Communications in Japan, A*, vol. 53, no. 1, pp. 36-43, 1970.
- [42] D. Youla, "On the factorization of rational matrices," *IRE Transactions on Information Theory*, vol. 7, no. 3, pp. 172-189, 1961.
- [43] C. Oară, "Constructive solutions to spectral and inner-outer factorizations with respect to the disk," *Automatica*, vol. 41, no. 11, pp. 1855-1866, 2005.
- [44] N. Bruinsma and M. Steinbuch, "A fast algorithm to compute the  $H_\infty$ -norm of a transfer function matrix," *Systems & Control Letters*, vol. 14, no. 4, pp. 287-293, 1990.
- [45] S. Boyd and V. Balakrishnan, "A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its  $L_\infty$ -norm," *Systems & Control Letters*, vol. 15, no. 1, pp. 1-7, 1990.
- [46] Y. Genin, P. Van Dooren, and V. Vermaut, "Convergence of the calculation of  $H_\infty$ -norm and related questions," in *Proceedings of MTNS*, 1998, pp. 429-432.
- [47] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. Prentice Hall New Jersey, 1996, vol. 40.
- [48] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Prentice Hall, 1978.
- [49] J. E. Markel and A. H. Gray, *Linear Prediction of Speech*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1982.
- [50] X. Huang, A. Acero, H.-W. Hon, and R. Reddy, *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR Upper Saddle River, 2001, vol. 95.
- [51] Public GitHub repository: <https://github.com/baggiogi/finsler-speech-morphing>.
- [52] A. Ferrante, A. Lanzon, and L. Ntogramatzidis, "Foundations of not necessarily rational negative imaginary systems theory: Relations between classes of negative imaginary and positive real systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 3052-3057, 2016.
- [53] —, "Discrete-time negative imaginary systems," *Automatica*, vol. 79, pp. 1 - 10, 2017.



**Giacomo Baggio** received the "Laurea Magistrale" degree (M. Sc. equivalent) and the Ph.D. degree in Control Systems Engineering from the University of Padova in 2014 and 2018. He is currently a PostDoctoral Scholar in the Department of Mechanical Engineering of the University of California at Riverside. From October 2015 to June 2016, he spent a visiting period at the Department of Engineering, University of Cambridge (UK). His research interests are in the areas of stochastic realization, spectral estimation, and control and analysis of dynamical networks.



**Augusto Ferrante** was born in Piove di Sacco, Italy, on August 5, 1967. He received the "Laurea" degree, *cum laude*, in Electrical Engineering in 1991 and the Ph.D. degree in Control Systems Engineering in 1995, both from the University of Padova.

He has been on the faculty of the Colleges of Engineering of the University of Udine, and of the "Politecnico di Milano". He is presently Professor in the "Department of Information Engineering" of the University of Padova.

His research interests are in the areas of linear systems, spectral estimation, optimal control and optimal filtering, quantum control, and stochastic realization.



**Rodolphe Sepulchre** (M'96-SM'08-F'10) received the engineering degree in 1990 and the Ph.D. degree in 1994, both in mathematical engineering, from the Universite Catholique de Louvain, Belgium. He was a BAEF Fellow in 1994 and held a postdoctoral position at the University of California, Santa Barbara from 1994 to 1996. He was a Research Associate of the FNRS at the Universite Catholique de Louvain from 1995 to 1997. Since 1997, he has been Professor in the Department of Electrical Engineering and Computer Science at the Universite de Liege.

He held visiting positions at Princeton University (2002-2003) and the Ecole des Mines de Paris (2009-2010) and part-time positions at the University of Louvain (2000-2011) and at INRIA Lille Europe (2012-2013). Since 2013, he is Professor of Engineering in the Department of Engineering, University of Cambridge, Cambridge, U.K. Dr. Sepulchre was awarded the IEEE Control Systems Society Antonio Ruberti Young Researcher Prize in 2008.