Geometric Methods for Spectral Analysis

Xianhua Jiang, Zhi-Quan (Tom) Luo, *Fellow, IEEE,* Tryphon T. Georgiou, *Fellow, IEEE*

***Abstract*—This paper explores a geometric framework for modeling non-stationary but slowly varying time series, based on the assumption that short-windowed power spectra capture their spectral character, and that energy transference in the frequency domain has a physical significance. The framework relies on certain notions of transportation distance and their respective geodesics to model possible non-parametric changes in the power spectral density with respect to time. We discuss the relevance of this framework to applications in spectral tracking, spectral averaging, and speech morphing.**

***Index Terms*—spectral metrics, geodesics, transportation distance, spectral analysis, spectral tracking, spectral averaging, speech morphing**

I. Introduction

Spectral analysis of time-series has been a remarkable tool across science and engineering for the best part of the past century. Historically, the modeling of time series has evolved in parallel with the birth of probability and stochastic process, blended with analytic function theory, orthogonal polynomials, Toeplitz operators, and now constitutes a mature discipline. Early on the subject received a strong impetus by the advent of the fast Fourier transform and subsequently, it was once again transformed by the influx of optimization methods along with notions of entropy, likelihood, and other statistical concepts. For the most part, the basic theoretic tools rely on the assumption that time-series are sample paths of wide-sense stationary, and possibly ergodic, processes [1]. Invariably, the key challenge to spectral estimation methods lies in trade-offs between resolution and variance that stem from the finiteness of the observation record. The challenge is further amplified when dealing with non-stationary time-series where various models and methods, mostly parametric, have been proposed to cope the changes in the power spectra over time. Evidently, such time-series are ubiquitous, from seismic recordings to speech, sonar, radar, etc. The present paper puts forth a non-parametric framework which is geometric in nature and aims to deal with modeling of non-stationary spectra.

Several methods have been proposed to deal with a slowly time-varying spectral content, and the existing approaches can be broadly classified into two categories. The first class is based on parametric estimation. For instance, Rao [2] proposed to approximate time-varying coefficients of an AR model by the first terms of a Taylor expansion while unknown parameters are being estimated via a weighted least-squares problem. Grenier [3] extended the well-known Levinson's AR model, Cadzow's ARMA model, Burg's method, and Prony's method to the time-varying case, under the restriction that the unknown coefficients are the weights of the linear combinations of known basis functions. Under the assumption that the coefficients of an ARMA model change slowly with time, Kaderli and Kayhan [4] obtained the AR coefficients from the time-varying modified Yule-Walker equations. The time-varying autocorrelations are estimated from the inverse Fourier transform of the evolutionary periodogram. Based on the estimated evolutionary cepstrum, they obtain the MA coefficients by a simple recursion. There have also been applications of Kalman filtering to tracking changes in the coefficients of suitable models under appropriate assumptions, see for instance [5, Sec. 3.3].

The second family of methods which are nonparametric, generally assume that the signal is stationary over sufficiently small time intervals, and carry out the spectral estimation in each interval, of which short time Fourier transform (STFT) is the most widely used method. STFT can be thought of as the inner product of the signal with a fixed window translating both in time and frequency. In a similar spirit, wavelet transforms evaluate the inner product of the signal with translated and dilated square integrable functions known as wavelets [6]. As a general extension of the classical Fourier analysis, time-frequency distributions (TFDs) approximate a non-stationary signal as the sum of several mono-component time-varying signals. For each such signal, a TFD tracks the spectral variation based on instantaneous frequency (IF), and provides the energy concentration of the signal around IF at each time [7]. Many well-known TFDs are generalized into a unified formulation, which is also called Cohen's bilinear class [8].

Several generalizations have been made to extend the concept of spectrum to non-stationary processes. One of the early attempts is the Page's instantaneous power spectrum, which is defined as the differentiation of the magnitude squared Fourier transform of the signal from the past to the current with respect to time [9]. Under the assumption that the non-stationary characteristics change slowly over time, Priestley [10] proposed the concept of evolutionary spectrum, which can be interpolated as the local energy distribution over frequencies. The evolutionary spectrum can be estimated by using the "double windows" technique.

In this paper, we propose geometric ways to model the evolution of the spectral content of slowly time-varying stochastic processes. More specifically, by regarding the short-windowed spectra as a path on the manifold of power spectral density functions, we can use notions of distances and geodesics to improve upon these estimations. To proceed, we first need a suitable metric to quantify the distance between two power spectral density functions. Typical distance measures that

This work was supported by the National Science Foundation, the Vincentine Hermes-Luh endowment, and the Digital Technology Center, University of Minnesota.

The authors are with the Department of Electrical & Computer Engineering, University of Minnesota, Minneapolis, MN 55455 {jiang082,luozq,tryphon}@umn.edu

have been introduced in various contexts, such as Kullback-Leibler divergence and Itakura-Saito distance, are not metrics in the rigorous sense. In recent years, certain suitable metrics were derived having a physical interpretation, one such is prediction metric [11] which is closely related to the Itakura-Saito distance and endows the space of admissible power spectral densities with a Riemannian structure. Naturally, metric geometry allows power spectral density functions to be thought of as point in a suitable space and provides a framework for problems of approximation, smoothing, averaging, etc. In the present paper, we focus on a certain transportation distance between power spectra. This has its roots in the Monge-Kantorovich problem of transportation of measures and has the added benefit that the metric is weakly continuous. This property is very important since it ensures that small changes in approximating, deforming, morphing power spectral densities correspond to small changes in their statistics.

The paper is organized as follows. Section II introduces certain alternative distance measures which are suitable for metrizing power spectra with the main focus on the transportation metric and its geodesics. We then develop optimization techniques for spectral tracking in Section III, and for spectral averaging in Section IV. Section V explores the use of such methods in applications, and we conclude with remarks for possibly future development in Section VI.

II. Distances and geodesics

A variety of distance measures have been proposed to quantify changes in power spectral densities [12], [13]. These have been linked to probabilistic concepts, statistical inference, as well as practical issues motivated by applications such as speech analysis. Among these, *Kullback-Leibler (KL) divergence* is firmly rooted in probability theory for its relevance in data compression, hypothesis testing, etc. For instance, it quantifies the degradation in source coding efficiency when a code is based on a particular probability distribution while the pdf of the source differs. It can be expressed as

$$d_{\mathrm{KL}}(f_0, f_1) := \int_{-\pi}^{\pi} f_0 \log \frac{f_0}{f_1} \frac{d\theta}{2\pi},$$

where f_0 and f_1 are *normalized* density functions with total integral being 1. When f_1 is an infinitesimal perturbation of f_0, the KL divergence gives rise to the *Fisher information metric*, i.e.,

$$d_{\mathrm{KL}}(f, f+\delta) \simeq \int_{-\pi}^{\pi} \frac{\delta^2}{f} \frac{d\theta}{2\pi}$$

by ignoring the higher order terms on δ [14]. This is a Riemannian metric, that is, one given by a quadratic form in the perturbation δ. Its induced geodesic can be expressed as

$$\sqrt{f_\tau} = \frac{\sin(1-\tau)\vartheta}{\sin\vartheta}\sqrt{f_0} + \frac{\sin\tau\vartheta}{\sin\vartheta}\sqrt{f_1} \qquad (1)$$

for $\tau \in [0,\ 1]$, and ϑ denotes the angle between $\sqrt{f_0}$ and $\sqrt{f_1}$, i.e.,

$$\cos\vartheta = \int_{-\pi}^{\pi} \sqrt{f_0 f_1} \frac{d\theta}{2\pi}.$$

The latter is also known as the *Bhattacharyya distance*. Geodesics in the Fisher metric in fact correspond to great circles on the "sphere" $\{\sqrt{f} \ : \ \int f = 1\}$ and geodesic distances to the corresponding arclength.

The KL divergence has been widely used in information theory, however it may not be the best choice in other applications. In particular, in the context of speech analysis and synthesis, the *Itakura-Saito distance* [15], [13] has been widely used to measure the mismatch between two spectra. This is given by

$$d_{\mathrm{IS}}(f_0, f_1) := \int_{-\pi}^{\pi} \left(\frac{f_0}{f_1} - \log \frac{f_0}{f_1} - 1 \right) \frac{d\theta}{2\pi}.$$

In this the density functions are not required to have the same mass. A gain-optimized expression of the Itakura-Saito distance, which is often called the *Itakura distance* [16], [13], has also been suggested and is given by

$$\begin{aligned} d_{\mathrm{I}}(f_0, f_1) &:= \min_{\alpha>0} d_{\mathrm{IS}}(f_0, \alpha f_1) \\ &= \log \int_{-\pi}^{\pi} \frac{f_0(\theta)}{f_1(\theta)} \frac{d\theta}{2\pi} - \int_{-\pi}^{\pi} \log \frac{f_0(\theta)}{f_1(\theta)} \frac{d\theta}{2\pi}. \end{aligned}$$

The exact same expression was obtained in [17] by considering the ratio of the "degraded" predictive error variance over the optimal variance by analogy to how the KL divergence quantifies degradation in source coding efficiency. When one spectral density is a perturbation of another, the Itakura distance leads (modulo a scaling factor of 2) to the Riemannian pseudo-metric [17]

$$d_{\mathrm{I}}(f, f+\delta) \simeq \int_{-\pi}^{\pi} \left(\frac{\delta(\theta)}{f(\theta)} \right)^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \frac{\delta(\theta)}{f(\theta)} \frac{d\theta}{2\pi} \right)^2 \qquad (2)$$

on density functions. This is analogous to the Fisher metric and we refer to it as the "prediction metric" due to its roots in quadratic prediction theory [17]. It turns out that geodesic paths f_τ ($\tau \in [0,1]$) connecting spectral densities $f_0,\ f_1$ and having minimal length can be explicitly computed as

$$f_\tau(\theta) = f_0^{1-\tau}(\theta) f_1^{\tau}(\theta), \qquad (3)$$

for $\tau \in [0,1]$, see [17]. Furthermore, the length along such "prediction geodesics" can be also *explicitly* computed in terms of the end points

$$d_{\mathrm{pg}}(f_0, f_1) = \sqrt{\int_{-\pi}^{\pi} \left(\log \frac{f_1(\theta)}{f_0(\theta)} \right)^2 \frac{d\theta}{2\pi} - \left(\int_{-\pi}^{\pi} \log \frac{f_1(\theta)}{f_0(\theta)} \frac{d\theta}{2\pi} \right)^2}, \qquad (4)$$

which is closely related to the L_2 norm of the difference of the log spectra used in speech processing [13].

While these geometric concepts have proven relevant in many applications, they lack a very important and natural property that is often desirable, weak continuity. By definition, a sequence of measures $\{\mu_k\}$ converges to μ weakly if for any continuous bounded function ϕ, $\int \phi d\mu_k \ \to \ \int \phi d\mu$ as $k \to \infty$. This is desirable for a notion of convergence of power spectra since small changes in the power spectral density ought to reflect small changes in the statistics and vice versa. For instance, a "proper" distance between two line-spectra ought to decrease as two lines move closer, but it is not the case for distances that fail to be weakly continuous. In order to bring in this additional property, in this paper, we consider a certain family of distances that have a long history in mathematics and relate to the so-called Monge-Kantorovich transportation problem [18, pg. 212].

In the context of "transportation" problem, density functions represent mass distributions and a cost is associated with transferring one unit of mass from one location to another. The idea is that by doing so, one gives a precise correspondence between two densities while the transportation distance quantifies the optimal cost for establishing this correspondence. The original formulation goes back to Gaspar Monge in 1781, while Leonid Kantorovich developed the modern formulation and theory after the second world war –interestingly, Leonid Kantorovich received the Nobel prize for related work and application of the theory in Economics.

For the purposes of this paper, we will only consider quadratic transference-cost. This is what is known as the *Wasserstein distance* of order 2. We do so due to the computational simplicity of the metric. Thus, for two density functions f_0 and f_1 supported on the space X (e.g., $X = [-\pi, \pi]$), the transportation distance is defined as follows:

$$d_{\mathrm{W}_2}(f_0, f_1) := \left\{ \inf_{T(x)} \int_X |x - T(x)|^2 f_0(x) dx \right\}^{\frac{1}{2}}$$

where, in general, $T(x)$ is a measure-preserving map between f_0 and f_1. That is,

$$|\det \nabla T(x)| f_1(T(x)) = f_0(x). \tag{5}$$

In the case that f_0 and f_1 are one dimensional power spectra, the transportation plan can be computed *explicitly* through

$$\int_{-\pi}^{\theta} f_0(\sigma) d\sigma = \int_{-\pi}^{T(\theta)} f_1(\sigma) d\sigma. \tag{6}$$

The map $T(\theta)$ satisfying (6) ensures the optimal transportation cost for moving f_0 to f_1. It allows determining $T(\theta)$ by integrating f_0, f_1 and comparing the respective values of their integrals [19, section 3.2]. The scaling function $T(\theta)$ is monotonically increasing since both f_0 and f_1 are positive.

When f_1 is a small perturbation of f_0, the transportation distance results in a Riemannian metric; see [18, section 7.6] where the more general high dimensional space is also considered. In the one dimensional case, this as well can be computed explicitly as

$$d^2_{\mathrm{W}_2}(f, f + \delta) \simeq \int_{-\pi}^{\pi} \frac{\Phi(\theta)^2}{f(\theta)} \frac{d\theta}{2\pi},$$

where

$$\Phi(\theta) = \int_{-\pi}^{\theta} \delta(\phi) \frac{d\phi}{2\pi}.$$

It is interesting to see how this metric may be seen as related to the Fisher information metric. Perturbations δ of the density function having integral zero, can be thought of as "tangent vectors" in the space of densities. The requirement that the integral is zero follows from the requirement that all densities have the same mass. Thus, in order to compute $d^2_{\mathrm{W}_2}(f, f+\delta)$ one needs to integrate this tangent vector δ and compute Φ –this is a linear operation. It is important to recall that f_0 and f_1 can be thought of as defined on the unit circle. Accordingly $\Phi(\theta)$ can be represented as the integral of δ starting from any point on $[-\pi, \pi]$, not necessary $-\pi$, and returning to it modulo 2π.

Geodesics f_τ ($\tau \in [0, 1]$) between two end "points" f_0 and f_1 are determined by a gradient flow [18, page 252], which, in this very special one-dimensional case, specifies the geodesic via

$$((1 - \tau) + \tau T'(\theta)) f_\tau((1 - \tau)\theta + \tau T(\theta)) = f_0(\theta), \tag{7}$$

for $T(\theta)$ computed from (6).

These geometric properties of spectral densities provide a variety of tools for spectral analysis. By regarding each spectral density as a point in this manifold, we can apply intuition from Euclidean geometry to spectral analysis. For instance, as will be discussed in the sequel, we can use a geodesic to fit density functions in complete analogy to Gauss's least-square line fitting. We can also ask for the mean, or the median of a set of estimated density functions. This type of processing holds the promise to provide additional versatile tools for fusing information from various sensors, modeling variability of spectra with time and space, etc. In particular, given two density functions, we may deform one into another by following the geodesic between them–an technique that can be used for speech morphing. These ideas will be developed in more detail in the following sections.

III. Spectral Tracking

We now focus on modeling non-stationary time series in the time-frequency domain through geodesic path fitting. Suppose we are given a sequence of power spectral densities

$$\mathcal{G} := \{g_{\tau_i}(\theta) \ : \ \theta \in [-\pi, \pi] \text{ for } i = 0, 1, \ldots, n\},$$

where τ_i is an increasing sequence of time-indices, normalized so that $\tau_0 = 0$ and $\tau_n = 1$. These power spectra may typically be obtained from time-series data using STFT, and τ_i ($i = 0, 1, \ldots, n$) may represent the mid-points, or another marker, of the corresponding time-windows. We propose to use the concept of the distance and geodesic of spectra to regularize the tracking of spectral density at different time intervals [20]. We will focus on the transportation metric and its geodesics to illustrate the idea. As usual, all spectra are normalized whenever the transportation metric or the corresponding geodesics are being considered.

Our objective is to determine a W_2-geodesic f_τ, $\tau \in [0, 1]$, which minimizes

$$J_{\mathcal{G}}(f_\tau) := \sum_{i=0}^{n} \left(d_{\mathrm{W}_2}(f_{\tau_i}, g_{\tau_i}) \right)^2.$$

Any geodesic f_τ is completely specified by two "points," in our case f_0, f_1. Alternatively, it is also specified by the transference plan T according to (6) and (7). The optimal choice of f_0, f_1, T needs to be determined from the data, i.e., the spectra $\mathcal{G}$ and the times τ_i ($i = 0, \ldots, n$). The aim of this section is to solve the corresponding optimization problem. This is done next.

Computation of $d_{\mathrm{W}_2}(f_{\tau_i}, g_{\tau_i})$ requires only the correspondence $\tilde{\theta} \mapsto \hat{\theta}$ for which

$$\int_{-\pi}^{\tilde{\theta}} g_{\tau_i}(\sigma) d\sigma = \int_{-\pi}^{\hat{\theta}} f_{\tau_i}(\sigma) d\sigma,$$

for $\tilde{\theta}, \hat{\theta} \in [-\pi, \pi]$. But,

$$\int_{-\pi}^{\tilde{\theta}} g_{\tau_i}(\sigma) d\sigma = \int_{-\pi}^{\hat{\theta}=(1-\tau_i)\theta+\tau_i T(\theta)} f_{\tau_i}(\sigma) d\sigma = \int_{-\pi}^{\theta} f_0(\sigma) d\sigma, \tag{8}$$

thus,

$$J_{\mathcal{G}}(f_\tau) = \sum_{i=0}^{n} \int_{-\pi}^{\pi} \left(\tilde{\theta} - ((1-\tau_i)\theta + \tau_i T(\theta))\right)^2 g_{\tau_i}(\tilde{\theta}) d\tilde{\theta}$$

where the correspondence

$$\tilde{\theta} \ \mapsto \ \theta \ \mapsto \ \hat{\theta} = ((1-\tau_i)\theta + \tau_i T(\theta))$$

can be unravelled from (8).

To simplify the above expression for $J_{\mathcal{G}}(f_\tau)$, we bring in the cumulative distribution functions

$$F(\theta) = \int_{-\pi}^{\theta} f(\sigma) d\sigma,$$

denoted by capital letters; that is, F_{τ_i} is the integral of f_{τ_i}, and similarly for g_{τ_i}. Then (8) reduces to

$$G_{\tau_i}(\tilde{\theta}) = F_{\tau_i}((1-\tau_i)\theta + \tau_i T(\theta)) = F_0(\theta) = F_1(T(\theta)). \tag{9}$$

Let $t = F_0(\theta)$, then $t \in [0,\ 1]$ and increases monotonically with respect to θ. We can represent θ by an inverse function of F, that is

$$\theta = F_0^{-1}(t), \quad \text{and similarly} \quad T(\theta) = F_1^{-1}(t), \quad \tilde{\theta}_i = G_{\tau_i}^{-1}(t).$$

In addition, t also equals to $G_{\tau_i}(\tilde{\theta})$ from (9). Thus

$$dt = dG_{\tau_i}(\tilde{\theta}) = g_{\tau_i}(\tilde{\theta}) d\tilde{\theta}.$$

Then the objective function can be re-written as

$$J_{\mathcal{G}}(f_\tau) \quad = \quad \sum_{i=0}^{n} \int_{0}^{1} \left(\tilde{\theta}_i - ((1-\tau_i)\theta + \tau_i T(\theta))\right)^2 dt$$

Since g_{τ_i}'s are given, $\tilde{\theta}_i$ is known or can be computed numerically, and the only unknowns are θ and $T(\theta)$ which are functions of t.

To solve this problem numerically, we can divide the range of t $[0,1]$, into N subintervals of equal length $1/N$ and denote by θ_k, $k = 0, 1, \ldots, N$ the values of θ for which

$$F_0(\theta_k) = \frac{k}{N}.$$

Similarly, let $\tilde{\theta}_{i,k}$ ($i = 0, \ldots, n,\ k = 0, 1, \ldots, N$) denote the values for which

$$G_{\tau_i}(\tilde{\theta}_{i,k}) = \frac{k}{N},$$

and by $\hat{\theta}_k$ the values for which

$$F_1(\hat{\theta}_k) = \frac{k}{N}.$$

Thereby, $J_{\mathcal{G}}(f_\tau)$ is approximated by the following finite sum

$$\mathbb{J} = \frac{1}{N} \sum_{i=0}^{n} \sum_{k=1}^{N} \left(\tilde{\theta}_{i,k} - \left((1-\tau_i)\theta_k + \tau_i \hat{\theta}_k\right)\right)^2. \tag{10}$$

The values of $\tilde{\theta}_{i,k}$ ($i = 0, \ldots, n,\ k = 0, 1, \ldots, N$) can be readily computed from the problem data $\mathcal{G}$, and the only unknowns in this "discretization" of $J_{\mathcal{G}}(f_\tau)$ are the vector of θ's, namely θ_k, $k = 0, 1, \ldots, N$ (which help determine F_0) and the vector of corresponding $\hat{\theta}$'s (which help determine F_1, and then T). Therefore, the spectral tracking problem can be solved numerically via the following convex quadratic program with linear constraints:

$$\begin{aligned} \min \{\mathbb{J} : \ \text{subject to } & -\pi \le \theta_k \le \theta_{k+1} \le \pi \\ \text{and } & -\pi \le \hat{\theta}_k \le \hat{\theta}_{k+1} \le \pi \\ & \text{for } 0 \le k \le N-1\}. \end{aligned} \tag{11}$$

The objective function is convex, so the optimal solution can be found efficiently [21].

If instead we use the prediction-geodesic distance in (4), the problem can still be formulated into a convex optimization problem. We aim to determine a prediction-geodesic f_τ, $\tau \in [0,1]$, which minimizes

$$J_{\mathcal{G}}(f_\tau) := \sum_{i=0}^{n} \left(d_{\mathrm{pg}}(f_{\tau_i}, g_{\tau_i})\right)^2.$$

Recall that the geodesic path f_{τ_i} between f_0 and f_1 belongs to the exponential family

$$f_{\tau_i}(\theta) = f_0^{1-\tau_i}(\theta) f_1^{\tau_i}(\theta).$$

If we discretize θ into N intervals from $-\pi$ to π, and let $\hat{f}_{\tau_i} = \log f_{\tau_i}$, $\hat{g}_{\tau_i} = \log g_{\tau_i}$, for the simplicity of the notation, we can work on the problem numerically by minimizing the following expression

$$\begin{aligned} \mathbb{J} = & \sum_{i=0}^{n} \sum_{k=1}^{N} (\hat{g}_{\tau_i}(k) - (1-\tau_i)\hat{f}_0(k) - \tau_i \hat{f}_1(k))^2 \\ & - \frac{1}{N} \sum_{i=0}^{n} (\sum_{k=1}^{N} (\hat{g}_{\tau_i}(k) - (1-\tau_i)\hat{f}_0(k) - \tau_i \hat{f}_1(k)))^2 \\ = & \sum_{i=0}^{n} (\hat{g}_{\tau_i} - (1-\tau_i)\hat{f}_0 - \tau_i \hat{f}_1)^{\mathrm{T}} (\mathbf{I} - \frac{1}{N} \mathbf{1}\mathbf{1}^{\mathrm{T}}) \\ & (\hat{g}_{\tau_i} - (1-\tau_i)\hat{f}_0 - \tau_i \hat{f}_1), \end{aligned} \tag{12}$$

where $\mathbf{I}$ is an $N \times N$ identity matrix, $\mathbf{1}$ is an $N \times 1$ vector with each component being 1. Since $(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}})$ is positive semi-definite, formulation (12) is a convex function. Note that because predictive metric is scale-invariant, the solution is not unique. To make the computation stable, one can add constraints on the norms of $\hat{f}_0$ and $\hat{f}_1$.

IV. Spectral Averaging

Constructing means, or averages, of a given data set has been of great interest for the purpose of modeling, smoothing, clustering, etc. For example in spectral analysis, the Welch estimator can be thought of as the arithmetic mean of the windowed periodograms [22], [23]. The arithmetic mean is often used to track the central tendency of the data. However in some applications, it may not be the best choice. For instance, if we have two Gaussian distributions with means μ_1 and μ_2, and the same variance σ^2, then their arithmetic mean is bimodal if $|\mu_1-\mu_2|>2\sigma$, and unimodal otherwise [24], while one would hope that the mean of two Gaussian distribution is still a Gaussian distribution. For this reason, we consider the transportation mean in this section.

The transportation mean arises naturally in the facility allocation. For example, one has several possible holes to fill each with probability w_i. Now one needs to allocate a pile of sand so that the cost of moving the sand to the possible holes will be minimized. Mathematically, it can be described in the following way. Given a set of nonnegative functions $f_i(\theta),\ i=1,\ \cdots,\ n,\ \theta\in[a,\ b]$ with equal total integrals being 1, the transportation mean is obtained through the following optimization problem:

$$\bar{f}=\arg\min_{f}\quad\sum_{i=1}^{n}w_i\left(d_{\mathrm{W}_2}(f,f_i)\right)^2, \tag{13}$$

where w_i's are weights with the sum 1. This problem has been studied in [25], [26], [27] for general multidimensional distributions. Below we shall provide the results for the one dimensional case where proofs are relatively straightforward.

Using the cumulative distribution functions, the problem in (13) is equivalent to

$$\begin{aligned}\bar{f}=\arg\min_{f}\quad&\sum_{i=1}^{n}w_i\int_a^b|\theta-T_i(\theta)|^2dF(\theta) \qquad (14)\\ \text{s.t.}\quad&F(\theta)=F_i(T_i(\theta)),\quad i=1,\ \cdots,\ n\end{aligned}$$

Let $t=F(\theta)$ and represent θ by an inverse function of F, that is

$$\theta=F^{-1}(t),\ \text{ and }\ T_i(\theta)=F_i^{-1}(F(\theta))=F_i^{-1}(t).$$

Since $f(\theta)$ is nonnegative, $F(\theta)$ must increase monotonically with respect to θ. Equivalently, $F^{-1}(t)$ has to increase monotonically with respect to t. In addition, $F^{-1}(t)$ has to satisfy the boundary condition, i.e., $F^{-1}(0)=a$ and $F^{-1}(1)=b$. Instead of solving the problem (14) directly, we can first seek the optimal F^{-1}. Defining $g(t)=F^{-1}(t)$, we consider the following optimization problem:

$$\begin{aligned}g^*(t)=\arg\min_{g(t)}\quad&\sum_{i=1}^{n}w_i\int_0^1|g(t)-F_i^{-1}(t)|^2dt \qquad (15)\\ \text{s.t.}\quad&g'(t)\geq 0,\ \forall\ t\in[0,\ 1]\\ &g(0)=a\\ &g(1)=b.\end{aligned}$$

After obtaining $g^*(t)$, we have the optimal cumulative distribution function $\bar{F}(g^*(t))=t$, and the optimal mean can be calculated as

$$\bar{f}(g^*(t))=1/\left(g^*(t)\right)'.$$

Theorem 1: Let

$$f_i(\theta)>0,\ \forall\ \theta\in[a,\ b],\ \text{ and }\int_a^b f_i(\theta)d\theta=1,$$

then the transportation mean $\bar{f}$ defined in (14) is unique and has an explicit form

$$\bar{F}(\sum_{i=1}^{n}w_iF_i^{-1}(t))=t$$

where $\bar{F}$ is the cumulative distribution function of $\bar{f}$, and $F_i^{-1}(t)$ is the inverse cumulative distribution function of f_i.

Proof: We will first solve the problem (15). Note that $f_i(\theta)$ is given, therefore $F_i^{-1}(t)$ is known or can be calculated numerically. Since $f_i(\theta)$ is nonnegative,

$$(F_i^{-1})'(t)\geq 0,\ \forall\ t\in[0,\ 1] \tag{16}$$

and

$$F_i^{-1}(0)=a,\qquad F_i^{-1}(1)=b. \tag{17}$$

If we ignore the constraints on $g(t)$, the unique minimizer is

$$\hat{g}^*(t)=\sum_{i=1}^{n}w_iF_i^{-1}(t). \tag{18}$$

Because of the conditions on $F_i^{-1}(t)$ in (16) and (17), $\hat{g}^*(t)$ satisfies

$$(\hat{g}^*)'(t)\geq 0,\ \forall t\in[0,\ 1],\qquad F_i^{-1}(0)=a,\ \text{ and }\ F_i^{-1}(1)=b$$

which are exactly the constraints in problem (15). Therefore, (18) is the optimal solution for (15). Consequently, the optimal transportation mean $\bar{f}$ in (14) can be obtained explicitly from

$$\bar{F}(\sum_{i=1}^{n}w_iF_i^{-1}(t))=t.$$

■

In the case where f_i's are delta functions

$$f_i(\theta)=\delta(\theta-\phi_i),$$

we have

$$F_i^{-1}(t)=\phi_i\ \text{for}\ 0<t<1.$$

Let $\bar{\phi}$ denote the arithmetic mean of $\{\phi_i\}$

$$\bar{\phi}=\sum_{i=1}^{n}w_i\phi_i.$$

Then the cumulative function $\bar{F}$ of the transportation mean satisfies

$$\bar{F}^{-1}(t)=\bar{\phi}\ \text{for}\ 0<t<1.$$

Consequently,

$$\bar{f}(\theta)=\delta(\theta-\bar{\phi}).$$

To compare the transportation mean with the arithmetic mean for Gaussian distributions, we have the following corollary:

Corollary 2: Let $\{f_i(\theta)=\mathcal{N}(\mu_i,\sigma_i^2),\ i=1,\ \cdots,\ n\}$, then their transportation mean $\bar{f}(\theta)=\mathcal{N}(\bar{\mu},\bar{\sigma}^2)$ where $\bar{\mu}$ and $\bar{\sigma}$ are the arithmetic means of $\{\mu_i\}$ and $\{\sigma_i\}$ respectively.

Proof: Let $f_0(\theta)=\mathcal{N}(0,1)$, then $f_i(\theta)$ is obtained by translating and dilating $f_0(\theta)$, that is

$$\sigma_if_i(\sigma_i\theta+\mu_i)=f_0(\theta),\quad i=1,\ \cdots,\ n.$$

Consequently,

$$F_i(\sigma_i\theta + \mu_i) = F_0(\theta), \quad i = 1, \ \cdots, \ n.$$

Using Theorem 1, the transportation mean satisfies

$$\bar{F}\left(\sum_{i=1}^{n} w_i(\sigma_i\theta + \mu_i)\right) = F_0(\theta),$$

which is the same as

$$\bar{F}\left((\sum_{i=1}^{n} w_i\sigma_i)\theta + \sum_{i=1}^{n} w_i\mu_i\right) = F_0(\theta).$$

Thus $\bar{f}(\theta)$ is still a Gaussian distribution with mean and variance being

$$\bar{\mu} = \sum_{i=1}^{n} w_i\mu_i, \quad \bar{\sigma} = \sum_{i=1}^{n} w_i\sigma_i,$$

which are the arithmetic means of $\{\mu_i\}$ and $\{\sigma_i\}$ respectively. ■

V. Applications

In this section, we shall provide some applications of the ideas developed before. For all numerical examples that follow, we used the software package SeDuMi, available through [28], to solve optimization problems.

A. Geodesic Fitting

As a numerical example, we generate time-series data by driving a time-varying system with unit-variance white noise and then superimposing white measurement noise with variance equal to 2. The time-varying system consists of a succession of (15^{th}-order) auto-regressive filters chosen to match the spectral character of a W_2-geodesic between an ideal power spectrum

$$f_{0,\text{ideal}}(\theta) = |\frac{1 - 0.5z^{-1} + 0.6z^{-2}}{1 + 0.8z^{-1} + 0.9z^{-2}}|^2_{z=e^{j\theta}}$$

and a final

$$f_{1,\text{ideal}}(\theta) = |\frac{1 + 0.5z^{-1} + 0.6z^{-2}}{1 - 0.8z^{-1} + 0.9z^{-2}}|^2_{z=e^{j\theta}}.$$

These are shown in Figure 1. STFT with a window of 128 points and an overlapping between successive windows by 64 points provides us with a collection of power spectral $\mathcal{G}$ as before.

Figure 3 shows the time-series data (in the first row) and then, below, it compares STFT power spectra ($g_{\tau_i}(\theta)$) in the second row with corresponding spectra obtained via a geodesic fit ($f_{\tau_i}(\theta)$). Figure 2 compares the spectrogram obtained by STFT and the one by the geodesic path fitting. It is clear that the geodesic path captures quite accurately the drift of power in the spectrum over time. Furthermore, the corresponding "frozen time" spectra $f_{\tau_i}(\theta)$ for $i = 0, \ldots, n$ appear to reproduce quite accurately the expected power distribution at the particular points in time. On the other hand, due to the small signal to noise ratio (SNR) the STFT seem quite unreliable.

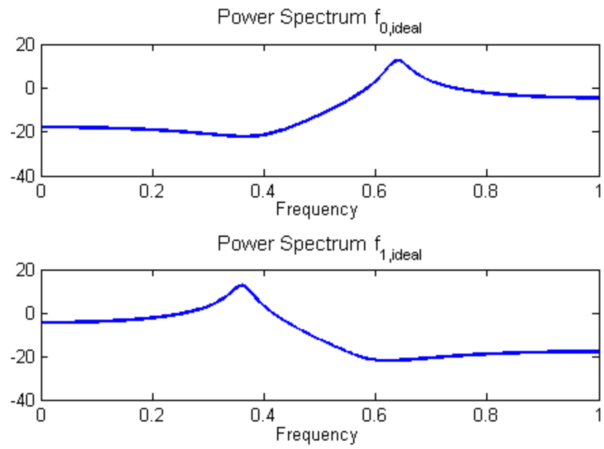


Fig. 1. Power spectra $f_{0,\text{ideal}}$ and $f_{1,\text{ideal}}$ as functions of θ/π.

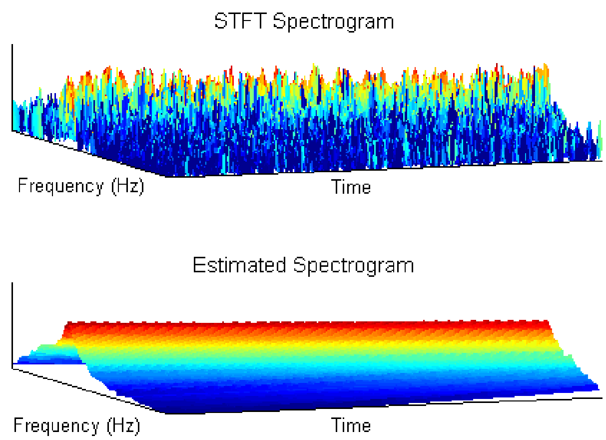


Fig. 2. STFT spectrogram and estimated geodesic path.

To deal with a long record of data where signal gradually undergoes significant changes, we can simply divide all spectra into several batches and determine a geodesic for each batch. However, this method can lead to discrepancies at the connecting points of successive batches. In order to ensure that no abrupt changes happen at these connecting points, we can add the distances between the ends of geodesics and the beginnings of the successive geodesics to the overall objective function. In addition, to guarantee that the signal changes slowly in each batch, we can also include all the length of geodesics as a penalty term. Therefore, our objective function becomes

$$J_{\mathcal{G}}(f_\tau) = \sum_{j=1}^{m}\sum_{i=0}^{n} (d_{\mathrm{W}_2}(f_{j,\tau_i}, g_{j,\tau_i}))^2 + \alpha \sum_{j=1}^{m-1} (d_{\mathrm{W}_2}(f_{j,\tau_n}, f_{j+1,\tau_1}))^2 + \beta \sum_{j=1}^{m} (d_{\mathrm{W}_2}(f_{j,\tau_1}, f_{j,\tau_n}))^2,$$

where j is the index of batch, m is the total number of batches, α and β are weights. In a similar way as before, $J_{\mathcal{G}}(f_\tau)$ can

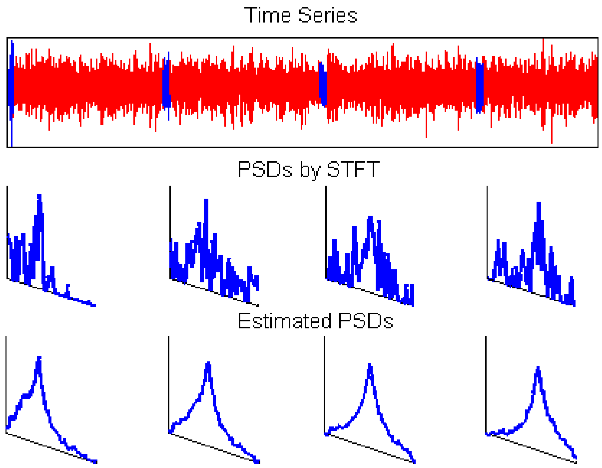


Fig. 3. First row: time-series data; second row: STFT spectra based on the highlighted parts of the time-series; third row: samples of geodesic fit to STFT spectra.

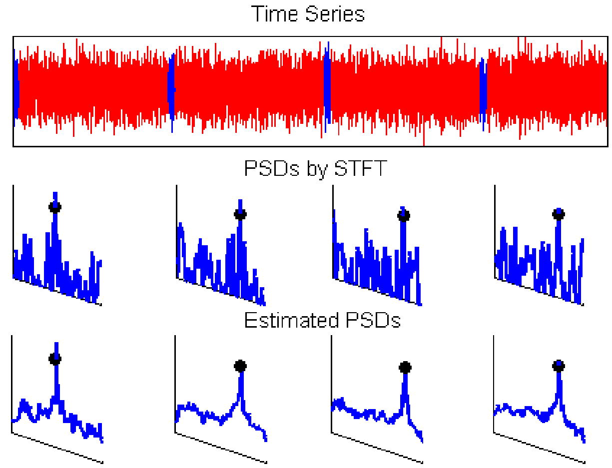


Fig. 5. Top row: time-series with two chirp signals and additive noise; second row: STFT spectra corresponding to windows marked with blue; third row: batched geodesic-fit samples without penalty terms.

be approximated as

$$\begin{aligned}\mathbb{J} =& \frac{1}{N}\sum_{j=1}^{m}\sum_{i=0}^{n}\sum_{k=1}^{N}\left(\tilde{\theta}_{j,i,k} - \left((1-\tau_i)\theta_{j,k} + \tau_i\hat{\theta}_{j,k}\right)\right)^2 \\ &+ \frac{\alpha}{N}\sum_{j=1}^{m-1}\sum_{k=1}^{N}\left(\hat{\theta}_{j,k} - \theta_{j+1,k}\right)^2 + \frac{\beta}{N}\sum_{j=1}^{m}\sum_{k=1}^{N}\left(\hat{\theta}_{j,k} - \theta_{j,k}\right)^2,\end{aligned}$$

where the definitions of $\tilde{\theta}_{j,i,k}$, $\theta_{j,k}$, and $\hat{\theta}_{j,k}$ are similar as before with j as the index of batch. The constraints resemble before by replacing θ_k and $\hat{\theta}_k$ with $\theta_{j,k}$ and $\hat{\theta}_{j,k}$. As an example, we generate a quadratic chirp signal with additive noise of standard variance 1. Figures 4 and 5 show the result with $\alpha = 50$, $\beta = 1$ and the batch size being 20. The black dots denote the true frequency of the signal at that time.

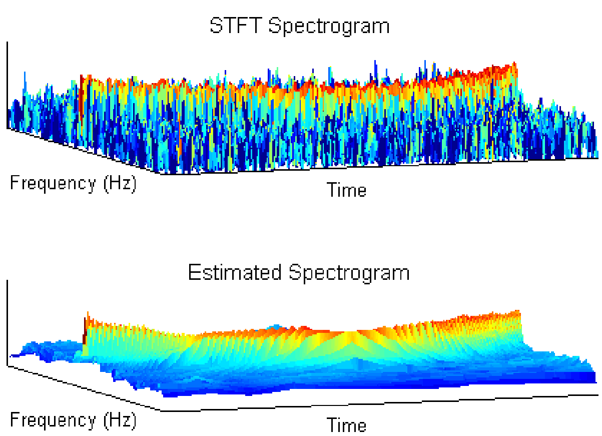


Fig. 4. STFT spectrogram and the batched geodesic path without penalty terms.

If the computational cost is not an issue, we can fit a geodesic for every pair of successive spectra so that the total geodesic length and the least square deviations from the given spectra are minimized. Therefore, our purpose is to determine f_{τ_i}, which minimizes

$$J_{\mathcal{G}}(f_\tau) := \sum_{i=0}^{n}\left(d_{\mathrm{W}_2}(f_{\tau_i}, g_{\tau_i})\right)^2 + \alpha\sum_{i=0}^{n-1}\left(d_{\mathrm{W}_2}(f_{\tau_i}, f_{\tau_{i+1}})\right)^2.$$

Note that in this case f_{τ_i}'s are not confined in a single geodesic. Let $F_{\tau_i}(\theta_{i,k}) = \frac{k}{N}$, where F_{τ_i} is the mass distribution function of f_{τ_i} as before, then the above objective function can be approximated numerically as

$$\mathbb{J} = \frac{1}{N}\sum_{i=0}^{n}\sum_{k=1}^{N}\left(\tilde{\theta}_{i,k} - \theta_{i,k}\right)^2 + \frac{\alpha}{N}\sum_{i=0}^{n-1}\sum_{k=1}^{N}\left(\theta_{i,k} - \theta_{i+1,k}\right)^2.$$

The unknown variables $\theta_{i,k}$ are subject to the constraints that they are non-decreasing with respect to k and their values are between 0 and 1. Since we have $n * N$ unknowns, it can be computationally expensive when the number of spectra n is large.

To illustrate the idea, we generate a chirp signal with the additive noise of standard variance 1.25. The total of 19 spectra by STFT and those estimated by pair-wise geodesic fitting are shown in Figure 6. Figure 7 compares several samples of spectra with the black dotes denote the true frequency of the signal at the corresponding time.

Comparing with the transportation distance, Figure 8 and 9 show the estimated spectra by using the prediction geodesic. We can see that the prediction geodesic results in a fade-in and fade-out affect on the spectral tracking, thus may not be a proper choice in this situation.

B. Spectral Averaging

The idea of the transportation mean can be applied to the smoothing of spectral densities. For a slowly time-varying signal, the spectral densities are estimated by short time Fourier transformation (STFT). At each "frozen" time, we use

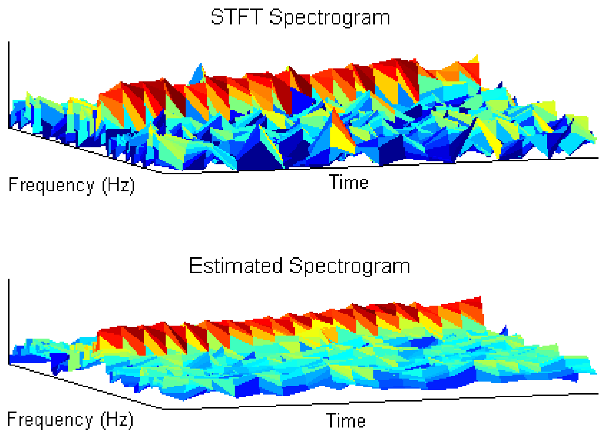


Fig. 6. STFT spectrogram and the pair-wise geodesic path.

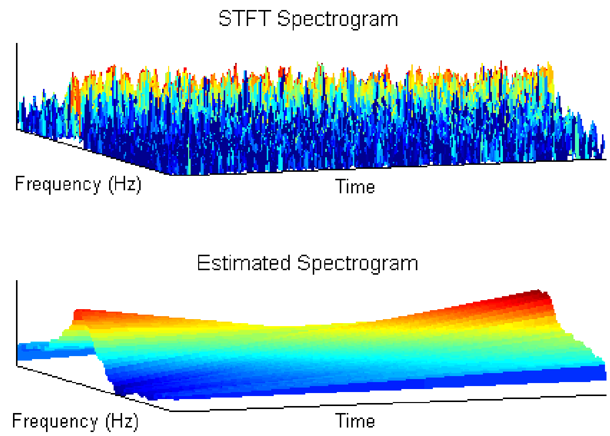


Fig. 8. STFT spectrogram and the prediction geodesic path.

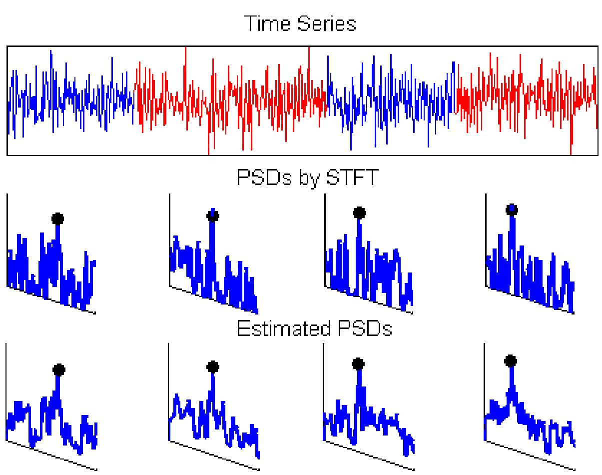


Fig. 7. Top row: time-series with two chirp signals and additive noise; second row: STFT spectra corresponding to windows marked with blue; third row: pair-wise geodesic-fit samples.

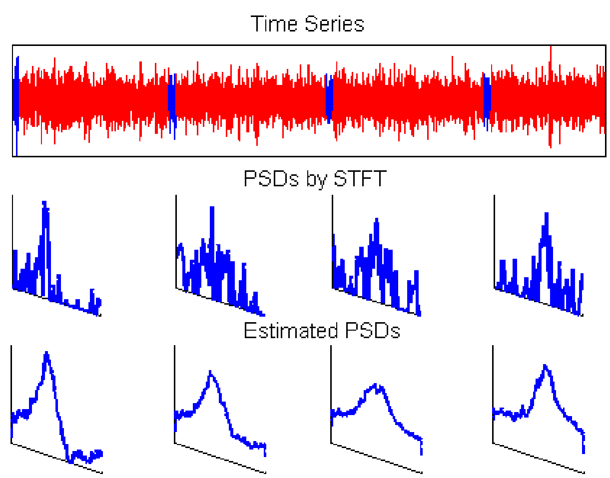


Fig. 9. First row: time-series data; second row: STFT spectra based on the highlighted parts of the time-series; third row: samples of prediction geodesic fit to STFT spectra.

the transportation mean of its spectral density and its neighbors to represent the smoothed spectral estimation.

As an example, we generated a chirp signal with additive noise. At each time, transportation mean is calculated with two neighbors from the past and two neighbors from the following with weights being $[0.1250\ 0.2188\ 0.3125\ 0.2188\ 0.1250]$. Figure 10 compares the spectral densities by STFT and the smoothed ones using the transportation mean. The black dots represents the true frequency at that time. We can see that spectral averaging has the potential of improving the resolution of the spectral estimation.

C. Speech Morphing

The purpose of this subsection is to consider geodesic paths between power spectral density functions as a means to morph voice of one individual to the voice of another [29]. Speech morphing is the process of transforming one person's speech pattern (e.g., Alice's) into another's (Bob's), gradually, creating a new pattern with a distinct identity, while preserving the speech-like quality and content of the spoken sentence. Despite great strides in the theory and technology of speech processing, speech morphing is still in an early phase and far from being a standard application [30], [31], [32], [33], [34], [30]. The key difference between our work and those earlier attempts is in the algorithm for interpolating the resulting power spectra. In our work we suggest that this can be most conveniently, and perhaps naturally, effected by following geodesics in suitable metrics. Below, we briefly discuss the steps taken for analyzing a voiced phoneme for A and B, generating the respective spectra, and then generating the morphed sound.

For analyzing voiced phonemes we use linear prediction techniques to obtain the coefficients for a modeling filter as well as to estimate the pitch. The frame size is 25 [ms], and the frame interval is 12 [ms]. A standard "pre-emphasis" filter is used to reduce the low-frequency content in the signal. The filtered data is weighted using a Hamming window. For modeling, we use the autocorrelation method for estimating the covariance lags and then the Levinson-Durbin method for obtaining the coefficients of the AR model. The AR model of the phoneme, for each of the two speaker, provides a corresponding power spectral density. A power spectrum at

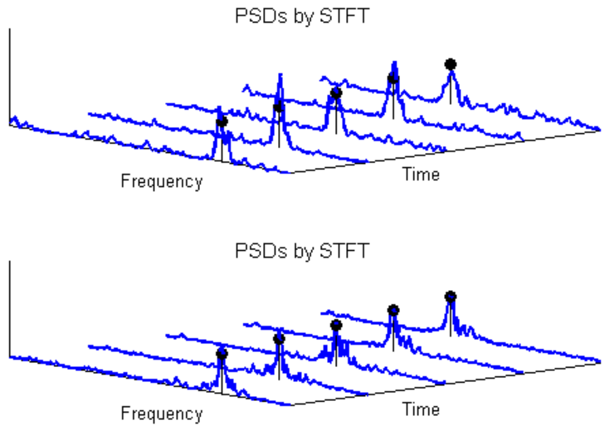


Fig. 10. First row: spectral densities by STFT; second row: smoothed spectral densities by transportation mean.

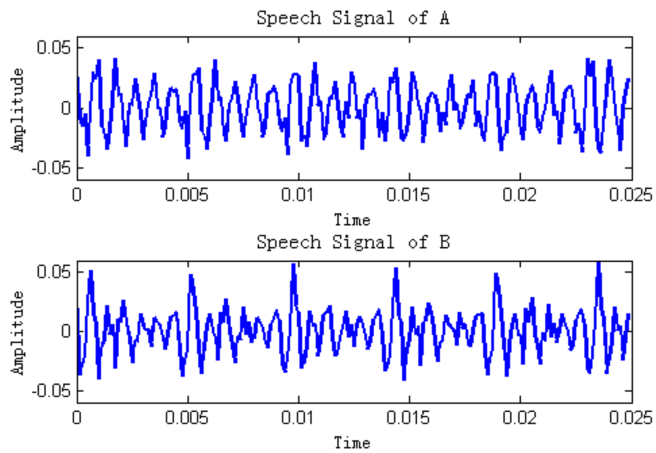


Fig. 11. Time signals corresponding to the phoneme "a" for A (female) and B (male)

a point $\tau \in [0, 1]$ on the geodesic path is then determined. We estimate the pitch period for speakers A and B. These periods are linearly interpolated along the path. The synthesized power spectral densities are approximated as AR-spectra and, finally, a pulse train with the interpolated pitch period is used to drive the synthesized AR-filter (at $\tau \in [0, 1]$ on the geodesic) in order to produce the synthesized sound. A post-emphasis filter is used to compensate the effect of the pre-emphasis filter–a standard practice. In synthesizing complete sentences, more complicated methods have been suggested to further improve the voice quality, such as a glottal-flow-like excitation [35]. With the purpose of illustrating the idea of spectral morphing, we didn't involve ourselves with these advanced techniques.

The voiced sound that is being analyzed (an "a") is shown as two superimposed time-waveforms in Figure 11 (one for speaker A and one for B); it was sampled at 8KHz. For modeling, we choose an AR model of order 14. The spectral deformation by the three alternative geodesics (1), (3), and (7) are shown in Figure 13, 14 and 15 respectively. Though the effect on the acoustic quality is very subjective, we find that geodesic (3) has surprisingly good acoustic qualities, in spite of the fact that visually, in Figures 14, there is an apparent "fade in" and "fade out" of the respective formants in the two power spectra. The voice quality by geodesic (7) has some artifacts, which may due to the gradually transport of formant locations and our ears may be very sensitive to it. The audio of speech morphing of the complete sentence "Thank you for your attention." using these techniques is included in a plenary presentation posted in [36]. The audio demonstrates intermediate reconstructions between the voices of two of the authors.

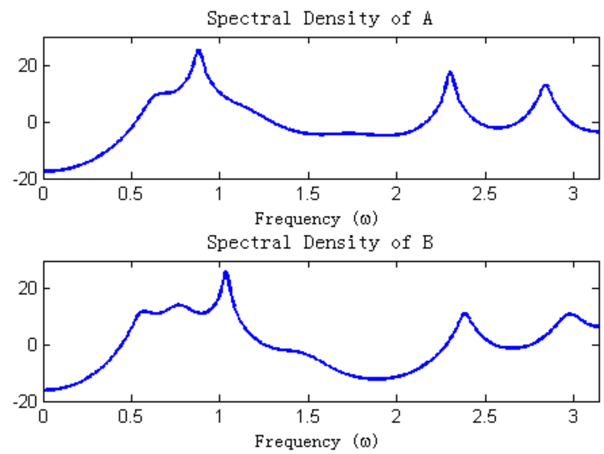


Fig. 12. Power spectra for subjects A (female) and B (male).

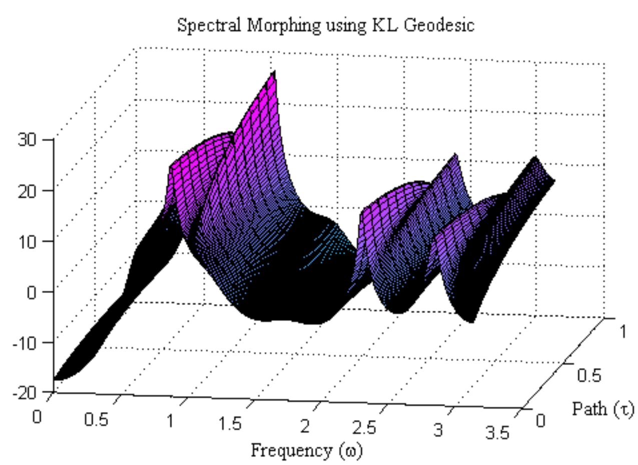


Fig. 13. Geodesic path between the two spectra following Fisher information metric.

VI. Conclusion

In recent years there has been an interest in endowing the space of power spectral densities with a natural metric (see [17], [37], [29], [38]). This has been motivated by a desire to develop tools for quantitative spectral analysis and modeling.

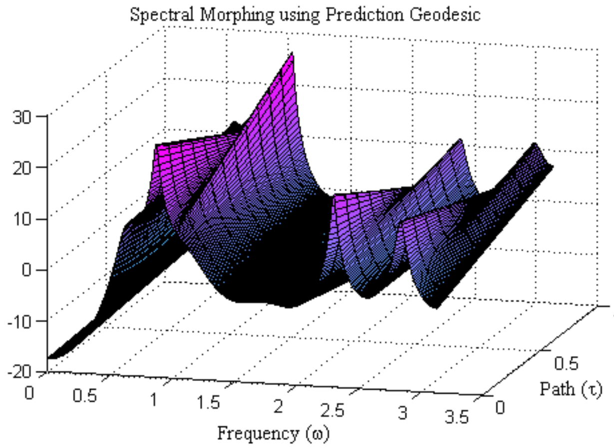


Fig. 14. Geodesic path between the two spectra following prediction metric.

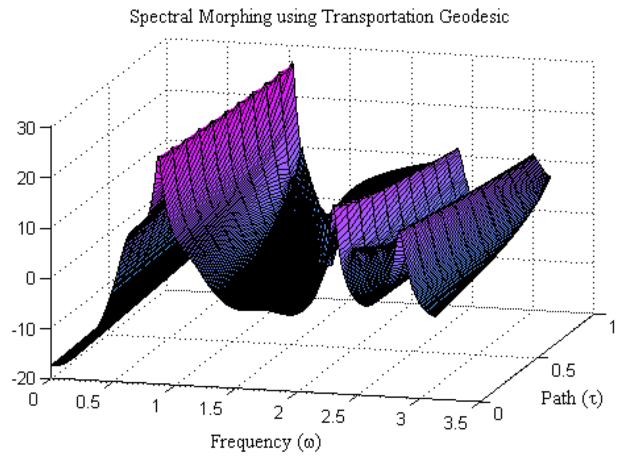


Fig. 15. Geodesic path between the two spectra following transportation metric.

Besides the relevance of metrics in quantifying modeling uncertainty, comparing spectra, etc., a metric topology brings up the concept of geodesics. These are the analogs of the straight lines of Euclidean geometry and represent the simplest models of paths across the space of density functions. In the present paper we sought to explore the concept of spectral geodesics for tracking features of a "time-varying" power spectrum. Such "time-varying spectra" are typically associated with non-stationary time-series [8], [39] modeling of which has always been somewhat of a conundrum in signal analysis. In this work, we formulated the problem of approximating spectra with geodesics utilizing the transportation metric. We then show that this "geodesic fit" problem is amenable to standard numerical tools of convex optimization. We use numerical examples to highlight the potential of the concept of a geodesic between spectra as a model for time-variability. The data for our formulations can be provided by standard spectral analysis techniques and, in particular, by the short-time Fourier transform. We also apply the idea of spectral geodesics to speech morphing, which could be potentially useful for voice transformation.

As a future direction, we expect to extend the transportation distance to distributions with unequal mass. Following earlier insights on this topic [40], [41], [42], [38], one may consider merging transportation distance with the L_2 distance. That is, at each step, for a spectral density f and its small perturbation $f+\delta$, we seek an intermediate point $\hat{f}$ such that the sum of the squared transportation distance between f and $\hat{f}$ and the squared L_2 distance between $\hat{f}$ and $f+\delta$ is minimized. We expect that the geodesic induced from such a mixed distance will provide a smooth deformation of both shape and total energy of the spectral density function which is natural from an applications' standpoint.

Acknowledgment

The authors would like to thank Daniel Rudoy for constructive discussions.

References

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[2] T. S. Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 32, no. 2, pp. 312–322, 1970.

[3] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 4, pp. 899–911, 1983.

[4] A. Kaderli and A. S. Kayhan, "Spectral estimation of nonstationary ARMA processes using the evolutionary cepstrum," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 130–132, 2002.

[5] B. Anderson and J. Moore, *Optimal Filtering*. Prentice-Hall, 1978.

[6] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[7] S. Haykin, *Advances in Spectrum Analysis and Array Processing (vol. I)*. Prentice Hall, 1991.

[8] L. Cohen, *Time-Frequency Analysis*. Prentice Hall, 1995.

[9] C. H. Page, "Instantaneous power spectra," *J. Appl. Phys.*, vol. 23, no. 1, pp. 103–106, 1952.

[10] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 27, no. 2, pp. 204–237, 1965.

[11] T. T. Georgiou, "Distances between power spectral densities," *IEEE Trans. Signal Process.*, vol. 55, pp. 3993–4003, 2007.

[12] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal process.*, vol. 18, no. 4, pp. 349–369, 1989.

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

[14] L. L. Campbell, "The relation between information theory and the differential geometry approach to statistics," *Inf. Sci.*, vol. 35, no. 3, pp. 199–210, 1985.

[15] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," *Proc. 6th Int. Congr. Acoust.*, pp. C17–C20.

[16] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 23, no. 1, pp. 67–72, 1975.

[17] T. T. Georgiou, "Distances and Riemannian metrics for spectral density functions," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 3995–4003, 2007.

[18] C. Villani, *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[19] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *Int. J. Comput. Vision*, vol. 60, no. 3, pp. 225–240, 2004.

[20] X. Jiang, Z. Luo, and T. T. Georgiou, "Power spectral geodesics and tracking," in *Proc. IEEE 47th Conference on Decision and Control (CDC'08)*, 2008, pp. 1315–1319.

[21] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[22] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, pp. 70–73, 1967.

[23] D. R. Brillinger, *Time Series: Data Analysis and Theory*. SIAM, 2001.

[24] J. Cohen, A. C. and P. J. Burke, "Compound normal distribution," *Amer. Math. Monthly*, vol. 63, no. 2, p. 129, 1956.

[25] M. Agueh and G. Carlier, "Barycenters in the wasserstein space," *SIAM J. Math. Anal.*, vol. 43, pp. 904–924, 2011.

[26] L. Rüchendorf and L. Uckelmann, "On the n-coupling problem," *J. Multivariate Anal.*, vol. 81, no. 2, pp. 242–258, 2002.

[27] M. Knott and C. S. Smith, "On the optimal mapping of distributions," *J. Optim. Theory Appl.*, vol. 43, no. 1, pp. 39–49, 1984.

[28] Sedumi software. [Online]. Available: http://sedumi.ie.lehigh.edu/

[29] X. Jiang, M. S. Takyar, and T. T. Georgiou, "Metrics and morphing of power spectra," *Recent Advances in Learning and Control*, pp. 125–135.

[30] H. R. Pfitzinger, "Unsupervised speech morphing between utterances of any speakers," in *Proc. 10th Australian International Conference on Speech Science and Technology (SST'04)*, 2004, pp. 545–550.

[31] M. Abe, "Speech morphing by gradually changing spectrum parameter and fundamental frequency," in *Proc. Fourth International Conference on Spoken Language Processing (ICSLP'96)*, 1996, pp. 2235–2238.

[32] H. Ye and S. Young, "Perceptually weighted linear transformations for voice conversion," in *Proc. 8th European Conference on Speech Communication and Technology (EUROSPEECH'03)*, 2003.

[33] H. Kawahara and H. Matsui, "Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. 1, 2003, pp. 256–259.

[34] V. Goncharoff and M. Kaine-Krolak, "Interpolation of LPC spectra via pole shifting," in *Proc. IEEEInternational Conference on Acoustics, Speech, and Signal Processing (ICASSP'95)*, vol. 1, 1995, pp. 780–783.

[35] M. R. Sambur, A. E. Rosenberg, L. R. Rabiner, and C. A. McGonegal, "On reducing the buzz in LPC synthesis," *J. Acoust. Soc. Am.*, vol. 63, no. 3, pp. 918–924, 1978.

[36] T. Georgiou. The meaning of distances in spectral analysis. Plenary lecture, IEEE Conference on Decision and Control, 2007. [Online]. Available: http://www.ieeecss-oll.org/video/meaning-distances-spectral-analysis

[37] J. Karlsson and T. T. Georgiou, "Signal analysis, moment problems & uncertainty measures," in *Proc. IEEE Conference on Decision and Control (CDC'05)*, 2005, pp. 5710–5715.

[38] T. T. Georgiou, J. Karlsson, and M. S. Takyar, "Metrics for power spectra: an axiomatic approach," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 859–867, 2009.

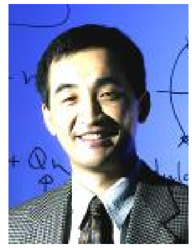
[39] P. Flandrin, *Time-frequency/Time-scale Analysis*. Academic Press, 1999.

[40] L. A. Caffarelli and R. J. McCann, "Free boundaries in optimal transport and Monge-Ampère obstacle problems," *Ann. Math. , to appear.*

[41] A. Figalli, "The optimal partial transport problem," *Arch. Ration. Mech. Anal.*, vol. 195, no. 2, pp. 533–560, 2010.

[42] J. D. Benamou and Y. Brenier, "Mixed L^2-Wasserstein optimal mapping between prescribed density functions," *J. Optim. Theory Appl.*, vol. 111, no. 2, pp. 255–271, 2001.

Xianhua Jiang received her B.Sc and M.Sc in Control Science and Engineering from the Zhejiang University, China in 2001 and 2004 respectively. From Sept. 2004 to May 2006, she was a research assistant on robotics and pattern recognition at the University of Vermont. She has been with the Department of Electrical and Computer Engineering at the University of Minnesota since Sept. 2006, where she is currently a Ph.D candidate. Her research interests lie in the areas of nonstationary signal analysis, speech processing, and spectral estimation.



Zhi-Quan (Tom) Luo (SM'03–F'07) received the B.Sc. degree in mathematics from Peking University, Peking, China, in 1984. From 1985 to 1989, he studied at the Operations Research Center and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, where he received the Ph.D. degree in operations research. During the academic year of 1984 to 1985, he was with the Nankai Institute of Mathematics, Tianjin, China. In 1989, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he became a Professor in 1998 and held the Canada Research Chair in Information Processing since 2001. Since April 2003, he has been a Professor with the Department of Electrical and Computer Engineering and holds an endowed ADC Research Chair in Wireless Telecommunications with the Digital Technology Center, University of Minnesota. His research interests lie in the union of large-scale optimization, information theory and coding, data communications, and signal processing.

Prof. Luo received an IEEE Signal Processing Society Best Paper Award in 2004. He is a member of the Society for Industrial and Applied Mathematics (SIAM) and MPS. He is also a member of the Signal Processing for Communications (SPCOM) and Signal Processing Theory and Methods (SPTM) Technical Committees of the IEEE Signal Processing Society. From 2000 to 2004, he served as an Associate Editor for the *IEEE Transactions on Signal Processing* and Mathematics of Computation. He is presently serving as an Associate Editor for several international journals including SIAM Journal on Optimization and Mathematics of Operations Research.

T. Georgiou (M'79–SM'99–F'00) was born in Athens, Greece, on October 18, 1956. He received the Diploma in Mechanical and Electrical Engineering from the National Technical University of Athens, Greece, in 1979, and the Ph.D. degree from the University of Florida, Gainesville, in 1983. He has served on the faculty of Florida Atlantic (1983-1986) and Iowa State (1986-1989) Universities. Since 1989 he has been with the University of Minnesota, Department of Electrical and Computer Engineering, where he is currently a Professor and holds the Vincentine Hermes-Luh Chair of Electrical Engineering. He has served as an associate editor for the IEEE Transactions on Automatic Control (1991-1992), the SIAM Journal on Control and Optimization (1988-1995), the Systems and Control Letters (1995-present), and the *IEEE Transactions on Signal Processing* (2007-2009). He has served on the Board of Governors of the Control Systems Society of the IEEE (2002-2005).

Dr. Georgiou has been a recipient of the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society three times, for the years 1992, 1999, and 2003. In 1992 and in 1999 he received the award for joint work with Prof. Malcolm C. Smith (Cambridge Univ., U.K.), and in 2003 for joint work with Professors Chris Byrnes (Washington Univ., St. Louis) and Anders Lindquist (KTH, Stockholm).