Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Minimal Itakura-Saito distance and Covariance interpolation

Per Enqvist and Johan Karlsson

Abstract— **Identification of power spectral densities rely on measured second order statistics such as, e.g. covariance estimates. In the family of power spectra consistent with such an estimate a representative spectra is singled out; examples of such choices are the Maximum entropy spectrum and the Correlogram. Here, we choose a prior spectral density to represent *a priori* information, and the spectrum closest to the prior in the Itakura-Saito distance is selected. It is known that this can be seen as the limit case when the cross-entropy principle is applied to a gaussian process.**

This work provides a quantitative measure of how close a finite covariance sequence is to a spectral density in the Itakura-Saito distance. It is given by a convex optimization problem and by considering its dual the structure of the optimal spectrum is obtained. Furthermore, it is shown that strong duality holds and that a covariance matching coercive spectral density always exists. The methods presented here provides tools for discrimination between power spectrum, identification of power spectrum, and for incorporating given data in this process.

I. Introduction

There has been a recent interest in distance measures for power spectra and in which way they discriminate. The choice of distance measure should reflect the application in mind, and this motivates the use of several different distance measures. Examples of this are the Kullback-Leibler ([1], [2], [3]) and the Hellinger distance ([4], [5]), which originated in probability theory, a set of distances quantifying the degradation of prediction [6], and metrics based on the Monge-Kantorovic transportation problem [7].

For identification and discrimination of power spectra the covariances are commonly used. For a finite number of covariances there is an infinite number of power spectra which match these second order moments exactly, and the spectrum is then chosen from this family of consistent estimates. Many methods for identification are based on this moment matching approach and how to make this choice defines the method. The so called central solution is based on finding the power spectrum that maximizes the entropy, which is equivalent to minimizing the Kullback-Leibler distance to the power spectrum corresponding to white noise.

If *a priori* information about the power spectrum is available then this can be used for choosing the spectrum from the family of consistent estimates. In particular, the set of degree constrained rational spectra consistent with the moments can be parameterized by using minimizers of the Kullback-Leibler distance to priors [3]. The problem of parameterizing all such solutions is known as the rational covariance extension problem.

Another measure that originates in probability theory is the Itakura-Saito ([8], [9]) distance. It was originally used in an approximative maximum likelihood derivation of the maximum entropy method, which is the solution corresponding to a flat prior. Using a more general prior will generate solutions different from the maximum entropy solutions. The minimization of the Itakura-Saito distance to the prior was advocated by Shore and Johnson as the cross-entropy principle; in [10] the cross-entropy principle is analyzed for probability densities and in [11] it is used for spectral analysis. A nice description of this approach is given in [12].

In this paper we revisit the minimization of the Itakura-Saito distance to an *a priori* power spectrum. This distance represents the difference in the asymptotic likelihood between the two spectra under gaussian assumptions. There are close connections between the Itakura-Saito distance, prediction error minimization and the recent advances on minimizing Kullback-Leibler distance in moment matching problems, see [13].

The problem of minimizing the Itakura-Saito distance to a prior subject to moment constraints is a convex infinite-dimensional problem. We consider its dual, which is finite-dimensional, and the structure of the optimal solution is derived. Furthermore, the solution of the dual gives the unique solution of the Itakura-Saito minimization. If the prior is rational then the minimizing spectrum is rational with degree bounded by the sum of the degree of the *a priori* spectrum and the number of estimated covariances.

One area of applications where it is important to be able to determine distances between spectral densities is in speech processing. Here we propose to use an Itakura-Saito distance with the known spectra of the phonems as an assumed prior knowledge and the covariances of the spectra to be classified for characterizing the unknown spectrum. This is decribed more in detail in Section VII. First, some concepts are defined in Section II, some distances are defined in Section III, the Itakura-Saito based moment matching problem is studied in Section IV, the relation to the Kullback-Leibler one in Section V, and the effect of scaling in Section VI. Then we finish with some conclusions in Section VIII and future work in Section IX.

II. Background

Let $(\ldots, y_{-1}, y_0, y_1, \ldots)$ be a stationary stochastic real-valued zero-mean process with covariances $r_k = \mathrm{E}\{\mathrm{y}_{\ell+\mathrm{k}}\mathrm{y}_\ell\}$ and power spectral density Φ. The power spectral density Φ

This work was supported by the Swedish research council.

The authors are with the division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden. penqvist@math.kth.se, johan.karlsson@math.kth.se

represents the energy content of the process across frequencies and has the covariances as Fourier coefficients,

$$\Phi(e^{i\theta}) = \sum_{k=-\infty}^{\infty} r_k e^{ik\theta}.$$

The covariance r_k is often estimated using the (biased) ergodic estimate

$$\frac{1}{N}\sum_{\ell=k+1}^{N} y_\ell y_{\ell-k}$$

based on a sample

$$(y_1, y_2, \cdots, y_N)$$

from the stationary process. This is a reliable estimate of the true covariances as long as $k \ll N$.

Consider the Hilbert space $L_2[-\pi, \pi]$ with the inner product

$$\langle a, b\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} a(e^{i\theta})b(e^{-i\theta})d\theta.$$

Then the covariances are given by

$$r_k = \left\langle \Phi, z^k \right\rangle.$$

Defining $\mathbf{z} = \begin{bmatrix} 1 & z & \cdots & z^n \end{bmatrix}^T$, the moments can be expressed as

$$\langle \Phi, \mathbf{z}\rangle = \mathbf{r},$$

where $\mathbf{r} = \begin{bmatrix} r_0 & r_1 & \dots & r_n \end{bmatrix}^T$, and the inner product is determined componentwise.

Given a covariance sequence $\mathbf{r}$, let $\mathfrak{F}_\mathbf{r}$ denote the set of power density functions consistent with $\mathbf{r}$, *i.e.*,

$$\mathfrak{F}_\mathbf{r} = \{\Phi \geq 0 | \langle \Phi, \mathbf{z}\rangle = \mathbf{r}\}.$$

Furthermore, we assume throughout that the symmetric Toeplitz matrix of the covariances

$$T(\mathbf{r}) = \begin{bmatrix} r_0 & r_1 & & r_n \\ r_1 & r_0 & \ddots & \\ & \ddots & \ddots & r_1 \\ r_n & & r_1 & r_0 \end{bmatrix}$$

is positive definite and hence the set $\mathfrak{F}_\mathbf{r}$ contains an infinite number of spectral densities.

III. Distance Measures

For two random variables, with sample space Ω, the Kullback-Leibler divergence is defined as

$$I_{\mathrm{KL}}(p, q) = \int_{x\in\Omega} p(x)\log\frac{p(x)}{q(x)}dx,$$

where the probability densities are p and q. This quantifies the average information for differentiating between the random variables. This is a non-negative measure, but it is not symmetric and it does not satisfy the triangle inequality. It has been frequently used within information theory and is also known as relative entropy, information gain, information divergence and directed divergence. Distance concepts for power spectral densities based on this probability density divergence are discussed next.

A. Kullback-Leibler Divergence

The Kullback-Leibler divergence was defined in [3] as a distance measure directly between spectral densities:

$$\mathbb{S}(\Psi||\Phi) \triangleq \left\langle \Psi, \log\frac{\Psi}{\Phi}\right\rangle. \tag{1}$$

The expression for $\mathbb{S}$ is analog to I_{KL}. Both probability densities and spectral densities are positive, but probability densities must have an integral equal to one. Positivity for $\mathbb{S}$ is ensured by the additional requirement on the spectral densities that $\int\Psi = \int\Phi$. This constraint can be satisfied by scaling the spectral densities and the distance then compares the shapes of the two spectra. A particularly interesting fact about $\mathbb{S}$ is that the rational covariance extension problem can be solved using Kullback-Leibler minimization [3].

B. Itakura-Saito distance

Here we define the Itakura-Saito distance for spectral densities Ψ and Φ such that the relevant integrals converge. Assume that Φ is coercive and Ψ and Φ both satisfy the Szegö condition, *i.e.*, there exists a positive ε such that $\Phi > \varepsilon$ on $[-\pi, \pi]$ and $\log\Phi$, $\log\Psi \in L_1[-\pi, \pi]$. Then,

$$D(\Phi, \Psi) = \left\langle \frac{\Phi}{\Psi}, 1\right\rangle - \left\langle \log\frac{\Phi}{\Psi}, 1\right\rangle - 1$$

is a distance measure that is non-negative and zero if and only if Φ and Ψ are the same. It should be noted that this measure is not symmetric with respect to the two arguments.

This distance measure between spectral densities can be derived by considering the divergence rate between two Gaussian vectors when the length of the vectors grows to infinity [8], [9], [14], [15].

In [8], also another representation of the distance is provided. If Ψ and Φ are close then the following sum will converge quickly

$$D(\Phi, \Psi) = \left\langle \frac{d^2(\lambda)}{2} - \frac{d^3(\lambda)}{3!} + \frac{d^4(\lambda)}{4!} - \ldots, 1\right\rangle,$$

where $d(\lambda) = \log\Psi - \log\Phi$ is the logarithmic difference between the two power spectral densities. This expression is used to show that this distance is more sensitive to peaks than to dips in the spectrum Φ.

The Itakura-Saito distance does not satisfy the triangle-inequality, so it is not a metric, but for some special (optimal) spectral densities it can be shown that the triangle-inequality do hold [10].

IV. Itakura-Saito Approximation with moment constraints

In moment matching based identification of spectral densities, specific choices of power spectra from the set $\mathfrak{F}_\mathbf{r}$ are made. Here we let an *a priori* spectral density Ψ determine the choice of spectrum taken from this set. Namely, we choose Φ as the spectra in $\mathfrak{F}_\mathbf{r}$ which is closest to Ψ in the Itakura-Saito distance. This distance provides a natural measure on how close an estimated sequence $\mathbf{r}$ is to the

spectrum Ψ, and can be used for identification and discrimination of stochastic processes. The spectral density Φ is thus determined by the following optimization problem.

$$\left[\begin{array}{ll} \min\limits_{\Phi} & \left\langle \frac{\Phi}{\Psi}, 1 \right\rangle - \left\langle \log \frac{\Phi}{\Psi}, 1 \right\rangle - 1 \\ \text{s.t.} & \Phi \in \mathfrak{F}_{\mathbf{r}}. \end{array}\right] \tag{2}$$

Note that here the spectral density is not constrained to be of a certain form, just as in the primal problem for Kullback-Leibler minimization, see [3] or (7) in Section V.

The optimization problem (2) has no finite dimensional parametrization, but by considering the dual, an optimization problem with a finite number of parameters is obtained.

Theorem 4.1: Let $\mathbf{r}$ be a positive covariance sequence and let Ψ be a rational positive function. The dual of the optimization problem (2) is given by

$$\left[\begin{array}{ll} \min\limits_{Q} & J(Q) = \langle R, Q\rangle - \langle 1, \log(1+Q\Psi)\rangle \\ \\ \text{s.t.} & 1 + Q(e^{i\theta})\Psi(e^{i\theta}) \geq 0, \quad \forall \theta. \end{array}\right] \tag{3}$$

where

$$Q(z) = \frac{1}{2}\sum_{k=0}^{n} q_k(z^k + z^{-k}), \tag{4}$$

and R is an arbitrary spectral density in $\mathfrak{F}_{\mathbf{r}}$. Furthermore, there is a unique solution Q to this problem such that $1 + Q(e^{i\theta})\Psi(e^{i\theta}) > 0$ for all $\theta \in (-\pi, \pi]$, and

$$\Phi = \Psi/(1+Q\Psi)$$

is the unique solution of the primal problem (2).

Remark 4.1: Another way to write the optimal form of the spectral density Φ is

$$\Phi = \frac{1}{\frac{1}{\Psi} + \frac{1}{1/Q}},$$

which is the harmonic mean of Ψ and $1/Q$, and corresponds to a "parallell" connection. But it should be noted that Q can be negative.

Proof: Form the Lagrangian function

$$L(\Phi; \mathbf{q}) \stackrel{\triangle}{=} \left\langle \frac{\Phi}{\Psi}, 1 \right\rangle - \left\langle \log \frac{\Phi}{\Psi}, 1 \right\rangle - 1 - \mathbf{q}^T\left(\mathbf{r} - \langle \Phi, \mathbf{z}\rangle\right)$$

where

$$\mathbf{q} = \left[\begin{array}{cccc} q_0 & q_1 & \cdots & q_n \end{array}\right]^T,$$

and since Φ is symmetric $\left\langle \Phi, \mathbf{q}^T\mathbf{z}\right\rangle = \langle \Phi, Q(z)\rangle$, where Q is defined in (4). Then the Lagrangian function can be written as

$$L(\Phi; \mathbf{q}) = \left\langle \frac{\Phi}{\Psi}, 1 \right\rangle - \left\langle \log \frac{\Phi}{\Psi}, 1 \right\rangle - 1 - \langle R, Q\rangle + \langle \Phi, Q\rangle.$$

Since L is strictly concave and differentiable with respect to Φ, the unique maximizing Φ is characterized by

$$\delta L = \left\langle \frac{1}{\Psi} - \frac{1}{\Phi} + Q, \delta\Phi \right\rangle = 0$$

for all $\delta\Phi$. Therefore, $\hat{\Phi} = \frac{\Psi}{1+\Psi Q}$. Then

$$L(\hat{\Phi}; \mathbf{q}) = -\langle R, Q\rangle + \langle 1, \log(1+\Psi Q)\rangle$$

where we used that

$$\left\langle \frac{1}{1+\Psi Q}, 1 \right\rangle + \left\langle \frac{\Psi}{1+\Psi Q}, Q \right\rangle = 1.$$

Maximizing L is equivalent to minimizing $-L$, so the dual (3) follows.

Recall that a direction d is called a *direction of recession* of a constraint $\mathcal{D}$ if for any $x \in \mathcal{D}$ it holds that $x + td \in \mathcal{D}$ for all $t > 0$. Let δQ be a direction of recession of the constraint $1 + Q(e^{i\theta})\Psi(e^{i\theta}) \geq 0$, then δQ is nonnegative and hence $\langle R, \delta Q\rangle > 0$. From this it follows that δQ is not a direction of recession of the function $J(Q)$, since

$$\sup_{Q}\left\{J(Q+\delta Q) - J(Q) \mid 1 + Q(e^{i\theta})\Psi(e^{i\theta}) \geq 0\right\} > 0.$$

Therefore, since J is a closed proper convex function and the domain is closed it follows from Theorem 27.3 in [16] that the dual problem attains a unique minimum.

The directional derivative of $J(Q)$ in the direction δQ is

$$\delta J(Q, \delta Q) = \left\langle \delta Q, R - \frac{\Psi}{1+Q\Psi} \right\rangle. \tag{5}$$

To show that the dual (3) has an interior optimal solution, we show that the derivative is infinite as we approach the boundary. Assume that the minimum is attained for a Q_0 such that $1+Q_0(e^{i\theta})\Psi(e^{i\theta}) \geq 0$ for all θ and $1+Q_0(e^{i\theta_0})\Psi(e^{i\theta_0}) = 0$ for some $\theta_0 \in (-\pi, \pi]$. Then, if $Q_t = Q_0 + t$ for $t > 0$, then $1+Q_t(e^{i\theta})\Psi(e^{i\theta}) > 0$ for all $\theta \in (-\pi, \pi]$. Now $\delta J(Q_t, D_t)$, where $D_t = (Q_0 - Q_t)/|Q_0 - Q_t| = -1$, grows to plus infinity as t goes to zero. Therefore, Q_0 can not minimize J and the optimal Q has to be in the interior. At the optimum the gradient (5) is zero, and hence $\Phi \in \mathfrak{F}_{\mathbf{r}}$ is the global optimum of (2). ■

A series of corollaries of Theorem 4.1 follows easily:

Corollary 4.1: If the prior spectral density $\Psi \in \mathfrak{F}_{\mathbf{r}}$, *i.e.* it is consistent with the covariances $\mathbf{r}$, then the optimal $Q = 0$ and $\Phi = \Psi$.

This is a simple consequence of the uniqueness of the solution.

A corollary observed by Politis in [17] is:

Corollary 4.2: If the prior spectral density is an all-pole model with at most n poles, *i.e.*, $\Psi = 1/A$ where the degree of the pseudo-polynomial A is at most n, then

$$\Phi = \frac{1}{A+Q}$$

is the spectral density of an all-pole model of degree n matching the covariances $r_0, r_1, \ldots, r_n$, *i.e.* it coincides with the Maximum Entropy solution.

Note that in this case there is a whole class of prior spectral densities such that the optimal spectral density is the same. In general, assume that Ψ_1 and Ψ_2 are two different priors generating the same interpolant Φ, *i.e.*

$$\frac{\Psi_1}{1+\Psi_1 Q_1} = \Phi = \frac{\Psi_2}{1+\Psi_2 Q_2}.$$

Then the relation

$$\Psi_2 = \frac{\Psi_1}{1+\Psi_1\tilde{Q}}, \quad \tilde{Q} = Q_1 - Q_2,$$

can be used to describe all priors generating the same spectral density, and extend Corollary 4.2.

Corollary 4.3: If Φ satisfies the moment constraints, then for all Q s.t. $1 - Q\Phi \geq 0$ the best approximant of

$$\Psi = \frac{\Phi}{1 - Q\Phi} \tag{6}$$

measured in the Itakura-Saito distance is Φ.

In particular, an arbitrary spectral density $\Phi \in \mathfrak{F}_{\mathbf{r}}$ is the optimal solution of an Itakura-Saito minimization problem (2) with the prior spectral density Ψ given by (6).

Corollary 4.4: If the prior spectral density is rational $\Psi = A/B$, then

$$\Phi = \frac{A}{B + AQ}$$

is also a rational spectral density, where the numerator is the same as for the prior and the denominator is of degree at most $\max\{n + \deg A, \deg B\}$, matching the covariances $r_0, r_1, \ldots, r_n$.

If A and B are not coprime, it follows easily also that neither A and $B + AQ$ will be coprime.

V. Relations between Kullback-Leibler and Itakura-Saito distance minimization

Consider minimizing the Kullback-Leibler distance $\mathbb{S}(\Psi||\Phi)$ under covariance interpolation constraints:

$$\left[\begin{array}{ll} \min\limits_{\Phi} & \left\langle \Psi, \log \dfrac{\Psi}{\Phi} \right\rangle \\ \text{s.t.} & \Phi \in \mathfrak{F}_{\mathbf{r}} \end{array} \right]. \tag{7}$$

In words; *Find the covariance matching spectral density of arbitrary form closest to the prior in the Kullback-Leibler distance.* This interpolation problem was solved using convex optimization methods in [18], and it is a special case of a problem studied in [3] where the Kullback-Leibler distance (1) to a prespecified spectral density Ψ is minimized.

Let $R \in \mathfrak{F}_{\mathbf{r}}$ be an arbitrary spectral density consistent with the given covariances. It has been shown in [18] that it is equivalent to solve the dual problem

$$\left[\begin{array}{ll} \min\limits_{Q} & \langle R, Q\rangle - \langle \Psi, \log Q\rangle \\ \\ \text{s.t.} & Q(e^{i\theta}) \geq 0, \quad \forall \theta. \end{array} \right] \tag{8}$$

where Q is defined as in (4). It can be observed that the dual objective function

$$\varphi_{\Psi}(Q) = \langle R, Q\rangle - \langle \Psi, \log Q\rangle$$

does not depend on which $R \in \mathfrak{F}_{\mathbf{r}}$ is chosen. Note also that it is equal up to a constant to

$$D_{\Psi}(R, \Phi) = \left\langle \frac{R}{\Phi}, 1 \right\rangle_{\Psi} - \left\langle \log \frac{R}{\Phi}, 1 \right\rangle_{\Psi} - 1$$

where $\Phi = \Psi/Q$. Here the inner products are evaluated in the measure $d\mu = \Psi\, d\theta$, so $\langle a, b\rangle_{\Psi} = \langle a, b\Psi\rangle$. This is a version of equation (18) in [13] where Remark 1 in the same paper is used to reformulate it. For $\Psi = 1$ it is clear that D_{Ψ} is the Itakura-Saito distance, and for a general Ψ it is a frequency weighted distance.

The dual problem (8) is then equivalent to

$$\left[\begin{array}{ll} \min\limits_{\Phi} & D_{\Psi}(R, \Phi) \\ \text{s.t.} & \Phi = \Psi/Q, \quad Q(e^{i\theta}) \geq 0, \quad \forall \theta. \end{array} \right]$$

In words; *Find the spectral density Φ of the form Ψ/Q closest to an arbitrary spectral density $R \in \mathfrak{F}_{\mathbf{r}}$ in the weighted Itakura-Saito distance.*

The weighting due to the prior appears also in the primal (7). Since Ψ is assumed to be fixed, it is equivalent to maximize $\langle \Psi, \log \Phi\rangle = \langle 1, \log \Phi\rangle_{\Psi}$ which can be seen as a frequency weighted entropy. In particular, it follows that for the prior $\Psi \equiv 1$, *i.e.*, the spectral density of a normalized white noise process, the maximum entropy solution is obtained (since it is the unique all-pole model in $\mathfrak{F}_{\mathbf{r}}$).

Also the dual of the Itakura-Saito minimization problem (3) can be seen to minimize an Itakura-Saito distance, by noting that

$$J(Q) = D(R, \Phi(Q)) - D(R, \Psi),$$

where $\Phi(Q) = \Psi/(1 + Q\Psi)$. So in (3) the Itakura-Saito distance to R is minimized over all spectral densities Φ of the special form $\Phi(Q)$.

VI. Scaling of the prior

One of the differences between the Kullback-Leibler and Itakura-Saito distances is that the former is only defined for spectra with same energy, $\int \Phi = \int \Psi$, and hence works more as a shape recognizer. For the Itakura-Saito distance, on the other hand, the variance of the spectra affects the distance directly. Therefore, it is interesting to consider the effect of scaling the prior spectrum Ψ in the minimization problem (2).

Let the prior spectrum be chosen as $\Psi_{\alpha} = \alpha\Psi_0$ for some positive Ψ_0 and $\alpha \in \mathbb{R}_+$. Given covariances $\mathbf{r}$ and the prior Ψ_{α}, the optimization problem (2) have a solution that will be called Φ_{α}. The objective function is modulo a constant term equal to

$$\int \frac{\Phi_{\alpha}}{\alpha\Psi_0} - \log(\Phi_{\alpha}).$$

The choice of α dictates the relative importance of the two terms in the cost. When α tends to infinity only the logarithmic term is important and the solution Φ_{α} will then tend to the maximum entropy solution. When α tends to zero only the first term is important. Minimizing $\langle \Phi/\Psi_0, 1\rangle$ is a linear programming problem in Φ. The dual of the limiting case is

$$\begin{array}{ll} \max\limits_{Q} & \langle R, Q\rangle \\ \text{s.t.} & 1 + Q\Psi_0 \geq 0 \end{array}$$

and in general the limiting "solution Φ_0" is represented by a measure with at most $2n$ point masses. Thus by letting α go from 0 to ∞ the resulting approximation go from a deterministic spectrum to the most unpredictable power spectrum, consistent with the covariance estimates.

The extreme cases corresponds to having large differences between the prior and the given covariance moments. The scaling of the prior can be used as a design parameter that regulate the determinism of the designed model. If the prior is scaled so that $\langle \Psi, 1\rangle = r_0$ an even weighting is obtained and the shape of the power spectrum will be at focus.

VII. Discrimination and matching using the Itakura-Saito distance

Speech processing is one area in which discrimination between power spectra is extensively used. Typically a set of features are determined on which the identification is performed. Commonly used features based on the spectra are cepstral coefficients, reflection coefficients or the spectrum itself. Here we will briefly discuss some aspects of this which relies on the Itakura-Saito distance for discrimination.

As a first step in discrimination, templates for different events needs to be determined. One way to do this is by using clustering algorithms such as the K-means algorithm [19, pp. 70], which focus on finding a set of templates to represent a large sample of spectra in a locally optimal way. For this, one part of the algorithm is to find centroids, i.e. a spectrum which minimize

$$\min_{\Psi} \sum_{\ell=1}^{m} D(\Psi, \Phi_\ell) \tag{9}$$

As the following proposition show, the minimizer of (9) is the harmonic mean.

Proposition 7.1: Let $\{\Phi_\ell\}_{\ell=1}^m$ be power spectral densities. Then

$$\min_{\Psi} \frac{1}{m} \sum_{\ell=1}^{m} D(\Psi, \Phi_\ell) = \int \log \frac{\left(\prod_{\ell=1}^m \Phi_\ell\right)^{\frac{1}{m}}}{\frac{m}{\sum_{\ell=1}^m \frac{1}{\Phi_\ell}}},$$

and the minimizer is

$$\Psi = \frac{m}{\sum_{\ell=1}^m \frac{1}{\Phi_\ell}}.$$

This shows that the accuracy with which a set of spectra could be approximated by one spectra depends on the ratio between the geometric and the harmonic means of the spectra. This and centroids for the Kullback-Leibler distance and the logarithmic 2-norm can be found in [20]. See [21] for more recent progress on centroids.

Once a set of power spectra $(\Psi_1, \Psi_2, \ldots, \Psi_p)$ representing a set of events are determined, Itakura-Saito approximation may be used to quantify how close a covariance sequence $\mathbf{r}$ is to the different template spectra. This is a straightforward procedure, but it illustrates the use of distances for discrimination.

VIII. Conclusions

We have investigated an approach for identification of power spectra subject to moment constraints using a priori information. The method relies on finding the spectrum closest to a given prior in the Itakura-Saito distance. This convex problem is reformulated to a finite-dimensional problem using duality theory for convex functions. The solution of this problem gives the desired spectrum and the minimal distance to the prior. We also study how the solution depends on scaling, a relation to Kullback-Leibler approximation, and discuss briefly how to utilize this for discrimination.

IX. Future works

Different distances can of course be used to find the best approximate in a specified class of systems. Using a global approach, where the spectral density is only assumed to satisfy some analytic properties, a duality argument can be used to show that the best approximant has to have a special structure. Recent work on the Hellinger distance [4], [5] has shown that for this distance the natural class of models is rational functions where the poles appears doubled. If Ψ is the prior power spectral density the structure of the best approximants in the Kullback-Leibler, Hellinger and Itakura-Saito distances are given by

$$\Phi_{\mathrm{KL}} = \frac{\Psi}{Q}, \quad \Phi_{\mathrm{H}} = \frac{\Psi}{(1+Q)^2}, \quad \Phi_{\mathrm{IS}} = \frac{\Psi}{1+Q\Psi}.$$

From this overview it is evident that for all these distances the optimal spectral density inherits the zeros of the prior spectral density. For the Kullback-Leibler and Hellinger distances any poles of the prior are also poles of the best approximations, Conversely, for the Itakura-Saito distance this is not the case.

One advantage with the Hellinger-based approach is that it generalize nicely to the MIMO case. It is well-known that the Itakura-Saito distance generalize to matrix valued power spectral densities.

$$\mathbf{D}(\Psi, \Phi) = \frac{1}{2} \left\langle \log \frac{\det \Phi}{\det \Psi}, 1 \right\rangle + \left\langle \mathrm{tr}\{\Phi^{-1}(\Psi - \Phi)\}, 1 \right\rangle$$

Generalizations to the multichannel case exists [22], [23] and the results derived in this paper are expected to work well also in that case.

References

[1] C.I. Byrnes, S.V. Gusev, and A. Lindquist, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM Review*, vol. 43, pp. 645–675, 2001.

[2] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. on Automatic Control*, , no. 46, pp. 822–839, June 2001.

[3] T.T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Transactions on Information Theory*, vol. 49, pp. 2910–2917, Nov 2003.

[4] M. Pavon and A. Ferrante, "On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities," *Automatic Control, IEEE Transactions on*, vol. 51, no. 4, pp. 639–644, April 2006.

[5] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger vs. Kullback-Leibler multivariable spectrum approximation," 2007.

[6] T.T. Georgiou, "Distances and riemannian metrics for spectral density functions," *IEEE Trans. signal processing*, vol. 55, no. 8, pp. 3995–4003, Aug. 2007.

[7] T. T. Georgiou, J. Karlsson, and M. S. Takyar, "Metrics for power spectra: an axiomatic approach," *IEEE Trans. on Signal Processing*, submitted 2007.

[8] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proc. 6th Int. Congr. Acoust. (Tokyo, Japan)*, 1968, pp. C17–C20.

[9] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral denstity and formant frequencies," *Electron. Commun. Japan*, vol. 53-A, pp. 36–43, 1970.

[10] J. Shore and R. Johnson, "Properties of cross-entropy minimization," *Information Theory, IEEE Transactions on*, vol. 27, no. 4, pp. 472–482, Jul 1981.

[11] J. Shore, "Minimum cross-entropy spectral analysis," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 2, pp. 230–237, Apr 1981.

[12] N. Wu, *The Maximum Entropy Method*, Springer-Verlag, 1997.

[13] A. Lindquist, *Modeling, Estimation and Control: Festschrift in honor of Giorgio Picci on the occasion of his sixty-fifth Birthday*, chapter Prediction-error approximation by convex optimization, Springer-Verlag, 2007.

[14] M. Pinsker, *Information and Information stability of random variables and processes*, Holden-Day, 1964.

[15] A. A. Stoorvogel and J. H. van Schuppen, *Identification, Adaptation, Learning*, vol. 153, chapter System Identification with Information Theoretic Criteria, pp. 289–338, Springer-Verlag, 1996, NATO ASI Series.

[16] R. T. Rockafellar, *Convex analysis*, Princeton, N.J., 1970.

[17] D.N. Politis, "Arma models, prewhitening, and minimum cross entropy," *Signal Processing, IEEE Transactions on*, vol. 41, no. 2, pp. 781–787, Feb 1993.

[18] C.I. Byrnes, S.V. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM Journal on Control and Optimization*, vol. 37, no. 1, pp. 211–229, 1999.

[19] J. Deller Jr., J. Hansen, and J. Proakis, *Discrete-time processing of speech signals*, IEEE Press, 2000.

[20] J. Karlsson, "Spectral estimation and distortion measures in speech processing," M.S. thesis, Optimization and Systems Theory, KTH, April 2003.

[21] F. Nielsen and R. Nock, "On the centroids of symmetrized Bregman divergences," arXiv:0711.3242v1, Nov. 2007.

[22] R. Johnson and J. Shore, "Minimum cross-entropy spectral analysis of multiple signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 3, pp. 574–582, Jun 1983.

[23] B. Musicus and R. Johnson, "Multichannel relative-entropy spectrum analysis," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 34, no. 3, pp. 554–564, Jun 1986.