Constrained spectrum approximation in the Hellinger distance

Augusto Ferrante, Michele Pavon and Federico Ramponi

Abstract— **We consider the Georgiou-Lindquist problem of approximating a spectral density function with spectra that are consistent with given state-covariance. Rather than the Kullback-Leibler pseudo-distance, however, we employ the Hellinger distance. We characterize the optimal solution and provide an iterative scheme for the Lagrange multiplier matrix that allows to solve numerically the dual problem.**

I. INTRODUCTION

In [1], Georgiou and Lindquist have studied the problem of approximating a (prior) spectral density when new information becomes available in the form of asymptotic state-covariance statistics. The approximation has been carried out in the Kullback-Leibler pseudo-metric. This study arose naturally in the frame of a broad program on generalized analytic interpolation and generalized moment problems with complexity constraint carried out over a number of years by Byrnes, Georgiou, Linquist and co-workers, see [8], [9], [10], [11], [21], [22]. This body of work originates from various important problems in the field such as the covariance extension problem, spectral estimation and H^∞ control, cf. [9], [21], [10], [2] and references therein.

The choice of a entropy type criterion was motivated in [1] by the following fact. While studying questions of existence and parametrization of solutions of general moment problems, the above mentioned authors realized that the two parametric families of measures that played a central role were critical points of entropy like functionals, see [3, p.3]. In the latter paper and in [4], homotopy like methods are proposed as an effective tool to solve a class of scalar and multidimensional generalized moment problems.

Very recently, Georgiou has investigated other distances for power spectra, [5], [6]. His motivation lies in prediction theory, where the optimal Wiener-Kolmogorov predictor does not depend on the L^1 norm of the spectrum. Considering the degradation of performance when an optimal filter of one stochastic process is employed to predict a different stochastic process naturally led him to formulate a distance between *rays* of spectral densities. Approximation with respect to the latter distance amounts to approximating the "shape" of a spectrum as it is sensible to pursue in several applications such as speech processing. Since the *a priori* and approximating spectrum do not necessarily have the same zero-th Fourier coefficient, the Kullback-Leibler approximation also amounts in many situations to approximating the shape of the *a priori* density, see [1, Section III].

In this paper, we take a different approach to constrained spectrum approximation borrowing from mathematical statistics another metric, namely the *Hellinger distance* [17], [18]. This approach appears to be more suitable to approximating a specific spectral density function rather than just approximating its shape. Indeed, as is shown in [13], the Hellinger distance is a *bona fide* distance that arises naturally as the minimum L^2 distance between *spectral factors* of the two spectra. Moreover, the variational analysis can be carried out in a straighforward manner *also in the multivariate case*.

The outline of the paper is as follows. In Section 2, we introduce the Hellinger distance between spectral density functions and formulate the corresponding constrained approximation problem. In Section 3, we derive the form of the optimal solution in terms of a Hermitian multiplier matrix. The following section is devoted to the dual problem. A matricial iterative algorithm to solve numerically the dual problem is introduced at the beginning of Section 5. We then present simulation results where the optimal solution in the Hellinger distance is compared to the optimal solution in the Kullback-Leibler sense.

II. CONSTRAINED SPECTRUM APPROXIMATION

We adopt the same notation as in [1]. Let $C_+(\mathbb{T})$ be the everywhere positive, continuous functions on the unit circle. We consider the rational transfer function[1]

$$G(z) = (I - zA)^{-1}B, \qquad A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times 1}, \quad (1)$$

where A is a stability matrix, i.e. has all its eigenvalues in the open unit disc, and (A, B) is a reachable pair. The latter property, as is well known, is equivalent to the controllability matrix $\mathcal{C}_n = [B, AB, A^2B, \ldots, A^{n-1}B]$ having full rank. Let $\Psi \in C_+(\mathbb{T})$ represent the *a priori* estimate of the spectrum of an underlying zero-mean, wide-sense stationary stochastic process $\{y(n), n \in \mathbb{Z}\}$. We consider the situation where new data become available in the form of an estimate Σ of the state covariance of the system

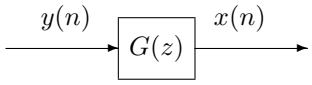
Work partially supported by the MIUR-PRIN Italian grant "Identification and Control of Industrial Systems"

A. Ferrante and F. Ramponi are with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy augusto@dei.unipd.it, rampo@dei.unipd.it

M. Pavon is with the Dipartimento di Matematica Pura e Applicata, Università di Padova, via Trieste 63, 35131 Padova, Italy pavon@math.unipd.it

[1]Notice that notation in (1) is different from the usual control engineering notation where z represents the forward shift. In this case, as it is customary in the mathematical literature, z represents the backward shift so that when A is a stability matrix, $G(z)$ is analytic inside (instead of outside) the unit disc.

ISBN: 978-960-89028-5-5



Here Σ, which is assumed to be positive definite, is obtained by feeding y to a bank of filters modeled by G until it reaches steady state, and then estimating the state covariance. When Ψ is not consistent with Σ, we need to find Φ in $C_+(\mathbb{T})$ that is closest to Ψ in a suitable sense among spectra consistent with Σ.

In order to turn this into a mathematical problem, we need to introduce a metric on spectral density functions. Given Φ and Ψ in $C_+(\mathbb{T})$, we define the *Hellinger distance* between them by

$$d_H(\Phi,\Psi) := \left[\int_{-\pi}^{\pi}\left(\sqrt{\Phi(e^{i\theta})}-\sqrt{\Psi(e^{i\theta})}\right)^2\frac{d\theta}{2\pi}\right]^{1/2}.$$

It is a *bona fide* distance on $C_+(\mathbb{T})$. Moreover, it satisfies the following properties.

Proposition 2.1: Consider $\Phi,\Psi\in C_+(\mathbb{T})$. Then

1) $d_H(\Phi,\Psi)\le\sqrt{\|\Phi\|_1+\|\Psi\|_1}$;
2) $d_H(\Phi,\Psi)^2\le\|\Phi-\Psi\|_1$;
3) $\|\Phi-\Psi\|_1\le\left(\sqrt{\|\Phi\|_1}+\sqrt{\|\Psi\|_1}\right)d_H(\Phi,\Psi)$.

Proof: Observe that

$$\int_{-\pi}^{\pi}\left(\sqrt{\Phi}-\sqrt{\Psi}\right)^2\frac{d\theta}{2\pi}=$$
$$\int_{-\pi}^{\pi}\left(\Phi+\Psi-2\sqrt{\Phi\Psi}\right)\frac{d\theta}{2\pi}\le\int_{-\pi}^{\pi}(\Phi+\Psi)\frac{d\theta}{2\pi}$$

which proves 1). Also

$$\int_{-\pi}^{\pi}|\Phi-\Psi|\frac{d\theta}{2\pi}=\int_{-\pi}^{\pi}|\sqrt{\Phi}-\sqrt{\Psi}||\sqrt{\Phi}+\sqrt{\Psi}|\frac{d\theta}{2\pi}\ge$$
$$\int_{-\pi}^{\pi}|\sqrt{\Phi}-\sqrt{\Psi}||\sqrt{\Phi}-\sqrt{\Psi}|\frac{d\theta}{2\pi}$$

which proves 2). Finally

$$\int_{-\pi}^{\pi}|\Phi-\Psi|\frac{d\theta}{2\pi}=\int_{-\pi}^{\pi}|\sqrt{\Phi}-\sqrt{\Psi}||\sqrt{\Phi}+\sqrt{\Psi}|\frac{d\theta}{2\pi}\le$$
$$d_H(\Phi,\Psi)\times\left[\int_{-\pi}^{\pi}\left(\Phi+\Psi+2\sqrt{\Phi\Psi}\right)\frac{d\theta}{2\pi}\right]^{1/2}\le$$
$$d_H(\Phi,\Psi)\left[\int_{-\pi}^{\pi}(\Phi+\Psi)\frac{d\theta}{2\pi}+2\sqrt{\int\Phi\int\Psi}\right]^{1/2}=$$
$$\left(\sqrt{\|\Phi\|_1}+\sqrt{\|\Psi\|_1}\right)d_H(\Phi,\Psi),$$

where we have used the Cauchy-Schwarz inequality twice. This establishes 3). Q.E.D.

We are now ready to formulate our approximation problem.

Problem 2.2: (**Approximation problem**) Let $\Psi\in C_+(\mathbb{T})$, and let $\Sigma\in\mathbb{C}^{n\times n}$ satisfy $\Sigma=\Sigma^*>0$. Find $\hat\Phi$ that solves

$$\text{minimize}\quad d_H^2(\Phi,\Psi) \tag{2}$$
$$\text{over}\quad\left\{\Phi\in C_+(\mathbb{T})\mid\int G\Phi G^*=\Sigma\right\}, \tag{3}$$

where star denotes transposition plus conjugation.

The constraint (3) expresses the fact that Φ is consistent with Σ. First of all, one needs to worry about existence of $\Phi\in C_+(\mathbb{T})$ satisfying constraint (3). It was shown in [21], [22] that such family is nonempty if and only if there exists $H\in\mathbb{C}^{1\times n}$ such that

$$\Sigma-A\Sigma A^*=BH+H^*B^*,$$

or, equivalently, the following rank condition holds

$$\text{rank}\begin{pmatrix}\Sigma-A\Sigma A^* & B\\ B^* & 0\end{pmatrix}=\text{rank}\begin{pmatrix}0 & B\\ B^* & 0\end{pmatrix} \tag{4}$$

It was shown in [1], that the constraint occurs in some important applications. In both examples below, the family of spectral densities consistent with the data is nonempty if $\Sigma\ge0$ and contains infinitely many elements if $\Sigma>0$.

Example 1: (*Covariance extension problem*) Let

$$A=\begin{bmatrix}0&1&0&\dots&0\\0&0&1&\dots&0\\\vdots&\vdots&&\ddots&\vdots\\0&0&0&\dots&1\\0&0&0&\dots&0\end{bmatrix},\quad B=\begin{bmatrix}0\\0\\\vdots\\0\\1\end{bmatrix} \tag{5}$$

so that the k-th component of $G(z)$ is $G_k(z)=z^{n-k}$. Also let Σ be the Toeplitz matrix:

$$\Sigma=\begin{bmatrix}c_0&c_1&\dots&c_{n-1}\\\bar c_1&c_0&\dots&c_{n-2}\\\vdots&\vdots&\ddots&\vdots\\\bar c_{n-1}&\bar c_{n-2}&\dots&c_0\end{bmatrix} \tag{6}$$

where

$$c_k:=E\{y(n)\bar y(n+k)\}.$$

Thus, the information available on the process y is the finite sequence of covariance lags $c_0,c_1,\dots,c_{n-1}$.

Example 2: In this case

$$A=\begin{bmatrix}p_1&0&0&\dots&0\\0&p_2&0&\dots&0\\\vdots&\vdots&&\ddots&\vdots\\0&0&0&\dots&0\\0&0&0&\dots&p_n\end{bmatrix},\quad B=\begin{bmatrix}1\\1\\\vdots\\1\\1\end{bmatrix} \tag{7}$$

so that the k-th element of $G(z)$ is $G_k(z)=\frac{1}{1-p_kz}$. The matrix Σ is a Pick matrix with elements $\Sigma_{i,j}=\frac{w_i+\bar w_j}{1-p_i\bar p_j}$ where

$$w_k=\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{e^{-i\theta}+p_k}{e^{-i\theta}-p_k}\Phi(e^{i\theta})d\theta,\quad k=1,2,\dots,n. \tag{8}$$

In this case the problem is a *Nevanlinna-Pick interpolation problem* [19], [20], [8], [9] whose solution permits *spectral estimation by selective harmonic amplification* [8], [9], [21].

III. VARIATIONAL ANALYSIS

The variational analysis is similar to the case when the Kullback-Leibler index is employed [1]. Define

$$\mathcal{L} := \{\Lambda \in \mathbb{C}^{n\times n} | \Lambda = \Lambda^*, 1 + G^*\Lambda G > 0 \; \forall e^{i\theta} \in \mathbb{T}\}.$$

For $\Lambda \in \mathcal{L}$, consider the *Lagrangian function*

$$\begin{aligned} L(\Phi, \Lambda) &= d_H^2(\Phi, \Psi) + \text{trace } \left(\Lambda \left(\int G\Phi G^* - \Sigma\right)\right) \\ &= d_H^2(\Phi, \Psi) + \int G^*\Lambda G\Phi - \text{trace } (\Lambda\Sigma). \quad (9) \end{aligned}$$

Next, consider the minimization of $L(\Phi, \Lambda)$:

$$\text{minimize}\{L(\Phi, \Lambda) | \Phi \in C_+(\mathbb{T})\}. \quad (10)$$

Remark 3.1: First of all, observe that $C_+(\mathbb{T})$ is an open, convex subset of $C(\mathbb{T})$. Second, notice that, for each Λ, the functional

$$\Phi \to L(\Phi, \Lambda)$$

is strictly convex. Thus, (10) is a convex optimization problem. Finally, observe that the smoothness of the integrands in (9) makes $L(\Phi, \Lambda)$ Gâteaux differentiable at Φ in any direction $\delta\Phi \in C(\mathbb{T})$.

It then follows from a basic result of convex optimization that $\hat{\Phi} \in C_+(\mathbb{T})$ solves problem (10) if and only if it satisfies the condition

$$\delta L(\hat{\Phi}, \Lambda; \delta\Phi) = 0, \quad \forall \delta\Phi \in C(\mathbb{T}). \quad (11)$$

Here, $\delta L(\hat{\Phi}, \Lambda; \delta\Phi)$, the first variation of L at $\hat{\Phi}$ in direction $\delta\Phi$, is defined by

$$\delta L(\hat{\Phi}, \Lambda; \delta\Phi) = \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[L(\hat{\Phi} + \epsilon\delta\Phi, \Lambda) - L(\hat{\Phi}, \Lambda)\right].$$

Proposition 3.2: The unique solution $\hat{\Phi}$ to problem (10) is given by

$$\hat{\Phi} = \frac{\Psi}{(1 + G^*\Lambda G)^2}. \quad (12)$$

Proof: For $\Phi \in C(\mathbb{T})$, we get

$$\delta L(\Phi, \Lambda; \delta\Phi) = \int \left(1 - \Psi^{1/2}\Phi^{-1/2} + G^*\Lambda G\right)\delta\Phi.$$

By (11), $\hat{\Phi} \in C_+(\mathbb{T})$ solves (10) if and only if

$$\int \left(1 - \Psi^{1/2}\Phi^{-1/2} + G^*\Lambda G\right)\delta\Phi = 0, \quad \forall \delta\Phi \in C(\mathbb{T}). \quad (13)$$

We get

$$\hat{\Phi}^{1/2} = \frac{\Psi^{1/2}}{1 + G^*\Lambda G}, \quad (14)$$

from which (12) follows. Q.E.D.

In the spirit of Lagrange, we get the following elementary, albeit fundamental, result.

Theorem 3.3: Suppose $\hat{\Lambda} = \hat{\Lambda}^*$ is such that

$$1 + G^*\hat{\Lambda}G > 0, \quad \forall e^{i\theta} \in \mathbb{T}, \quad (15)$$

$$\int G\frac{\Psi}{(1 + G^*\hat{\Lambda}G)^2}G^* = \Sigma. \quad (16)$$

Then $\hat{\Phi}$ given by

$$\hat{\Phi} = \frac{\Psi}{(1 + G^*\hat{\Lambda}G)^2} \quad (17)$$

is the unique solution of the approximation problem (2)-(3).

Proof: Let $\Phi \in C_+(\mathbb{T})$ satisfy the constraint

$$\int G\Phi G^* = \Sigma. \quad (18)$$

By Proposition (3.2), and by the strict convexity of the functional $L(\cdot, \hat{\Lambda})$, we get

$$d_H^2(\Phi, \Psi) = L(\Phi, \hat{\Lambda}) > L(\hat{\Phi}, \hat{\Lambda}) = d_H^2(\hat{\Phi}, \Psi).$$

By (16), $\hat{\Phi}$ in (17) satisfies the constraint (18). Hence it is optimal for the original constrained problem. Q.E.D.

Thus, the original problem (2)-(3) is now reduced to finding $\hat{\Lambda}$ satisfying (15)-(16). This is accomplished, as in [1], via duality theory.

IV. THE DUAL PROBLEM

To simplify the writing, we assume from now on that $\Sigma = I$. Indeed, if $\Sigma \neq I$, it suffices to replace G by $G' := \Sigma^{-1/2}G$ and (A, B) with $(A' = \Sigma^{-1/2}A\Sigma^{1/2}, B' = \Sigma^{-1/2}B)$. Moreover, if $\hat{\Lambda}'$ has been found such that

$$\int G'\frac{\Psi}{(1 + G'^*\hat{\Lambda}'G')^2}G'^* = I,$$

then

$$\int G\frac{\Psi}{(1 + G'^*\hat{\Lambda}'G')^2}G^* = \Sigma.$$

Hence, $\hat{\Phi}$ in (17) may also be obtained directly from G' and $\hat{\Lambda}'$ as

$$\hat{\Phi} = \frac{\Psi}{(1 + G'^*\hat{\Lambda}'G')^2}. \quad (19)$$

In view of Proposition 3.2, for Λ satisfying (15), the dual functional takes the form

$$\begin{aligned} &\Lambda \to L(\frac{\Psi}{(1 + G^*\Lambda G)^2}, \Lambda) = \\ &\int \frac{\Psi G^*\Lambda G}{1 + G^*\Lambda G} - \text{trace } (\Lambda). \quad (20) \end{aligned}$$

We consider the maximization of the dual functional (20) over the set $\mathcal{L} = \{\Lambda = \Lambda^* | 1 + G^*\Lambda G > 0, \forall e^{i\theta} \in \mathbb{T}\}$. Let

$$J_\Psi(\Lambda) := \text{trace } (\Lambda) - \int \frac{\Psi G^*\Lambda G}{1 + G^*\Lambda G}$$

The dual problem is then equivalent to

$$\text{minimize} \quad \{J_\Psi(\Lambda) | \Lambda \in \mathcal{L}\}. \quad (21)$$

Remark 4.1: Notice that $\mathcal{L}$ is convex. Moreover, $\Lambda \to J_\Psi(\Lambda)$ is convex on $\mathcal{L}$, but, in general, *not strictly convex*. More precisely, J_Ψ is strictly convex when restricted to the range of the operator Γ defined on $C(\mathbb{T})$ by

$$\Gamma(\Phi) = \int G\Phi G^*. \quad (22)$$

This can be established along the lines of [1, Section V]. Details will be provided in [13]. Observe that $\mathcal{L}$ is an open subset of the set $\mathcal{H}$ of all $n \times n$ Hermitian matrices and J_Ψ is Gâteaux differentiable in any direction $\delta\Lambda \in \mathcal{H}$.
Hence, $\hat{\Lambda} \in \mathcal{L}$ solves (21) if and only if, for all $\delta\Lambda \in \mathcal{H}$, we have

$$\delta J_\psi(\hat{\Lambda}; \delta\Lambda) = \text{trace} \left[\left(I - \int \frac{G\Psi G^*}{(1+G^*\hat{\Lambda}G)^2} \right) \delta\Lambda \right] = 0. \tag{23}$$

Arguing as in the minimization of the Lagrangian, we get the following result.

Proposition 4.2: $\hat{\Lambda} \in \mathcal{L}$ solves (21) if and only if it satisfies (16) with $\Sigma = I$.

Remark 4.3: Existence of a minimum of J_Ψ on $\mathcal{L}$ is established in [13] along the lines of [12] (a detailed existence proof for the Kullback-Leibler case is contained in [14]). This result implies that, under assumption (4), there exists a $\hat{\Lambda}$ in $\mathcal{L}$ satisfying (16). Such a $\hat{\Lambda}$ then provides the optimal solution of the primal problem (2)-(3) via (17). This, of course, establishes existence also for the dual problem (21) in view of Proposition (4.2).

As for the Kullback-Leibler case, a closed form solution of the dual problem may be obtained only in certain specific cases. In general, one needs to resort to an iterative scheme. We provide a simple method for the numerical solution of the dual problem in the next section.

V. SIMULATION RESULTS

Reparametrization of $\mathcal{L}$ or of a subset of it where to look for a solution to the dual problem has the advantage of leading to a vectorial iteration. It has the disadvantage, however, that it may lead to loss of convexity or unboundness of the gradient at the boundary, see [9], [23], [24] . As in [7] for the Kullback-Leibler case, we prefer therefore a matricial iterative scheme. In view of (23), one can easily recover an expression for the gradient of J_Ψ at Λ:

$$\nabla J_\Psi(\Lambda) = I - \int \frac{G\Psi G^*}{(1+G^*\Lambda G)^2}$$

and resort to a steepest descent-type algorithm. Simulations were performed using the simple iteration with constant step size $\alpha > 0$

$$\Lambda_{k+1} = \Lambda_k - \alpha \nabla J_\Psi(\Lambda_k), \quad \Lambda_0 = 0. \tag{24}$$

As observed in Remark 4.1, J_Ψ is strictly convex on $\text{Range}\,\Gamma$. By Remark 4.3, the identity matrix, and consequently $\nabla J_\Psi(\Lambda)$, belongs to $\text{Range}\,\Gamma$. Thus, if we start the iteration in $\text{Range}\,\Gamma$, Λ_k evolves in $\text{Range}\,\Gamma$. Local convergence of (24) can be established for a *sufficiently small* step size. As is well known, convergence in this method may be rather slow for bad conditioned problems. In that case, more efficient methods such as the backtracking line search descent described below (or a Newton-type method) may be employed [13]. For the backtracking line search descent, let $t = 1$, choose $0 < \beta < 1$ and $0 < \alpha < 1/2$ and consider the following iteration:

- While $J_\Psi(\Lambda_k - t\nabla J_\Psi(\Lambda_k)) > J_\Psi(\Lambda_k) - \alpha t ||\nabla J_\Psi(\Lambda_k)||^2$ *or* $\Lambda_k - t\nabla J_\Psi(\Lambda_k)$ does not belong to $\mathcal{L}$, decrease t by setting $t \leftarrow \beta t$;
- set $\Lambda_{k+1} := \Lambda_k - t\nabla J_\Psi(\Lambda_k)$.

Since it is a *descent* algorithm, the iteration evolves in the subset $S \cap \text{Range}\,\Gamma$, where $S = \{\Lambda = \Lambda^* | J_\Psi(\Lambda) \leq J_\Psi(\Lambda_0)\}$, whenever $\Lambda_0 \in \text{Range}\,\Gamma$. This algorithm is of the type "steepest descent with *backtracking* line search", see e.g. [25, Chapter 9].

Simulation 1. Consider the following instance of Example 1 (corresponding to a covariance extension problem):

$$A := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

and

$$\Sigma = I$$

Let $\Psi(z) = W_\psi(z) W_\psi^*(z)$, where

$$W_\psi(z) := K \frac{z - 1/2}{\left(z^2 + \frac{5}{100}z + 1/4\right)(z + 1/4)}.$$

Select $K = 0.7747$ so that $\int \Psi = 1$. We have iterated the algorithm (24) with the stopping condition $\|M - \Theta(M)\| < 10^{-9}$. The algorithm converged to the matrix

$$\hat{\Lambda}' = \begin{bmatrix} -0.0425 & -0.3160 \\ -0.3160 & -0.0425 \end{bmatrix}.$$

The optimal solution $\Phi^o(e^{i\omega})$ is depicted (bold line) together with $\Psi(e^{i\omega})$ (dashed line) in Figure 1. For the purpose of comparison, the optimal Kullback-Leibler approximant Φ^o_{KL} of [1] is also depicted in Figure 1 (solid line). In this case, the Hellinger approximant appears to be considerably better at low frequencies than the Kullback-Liebler approximant.

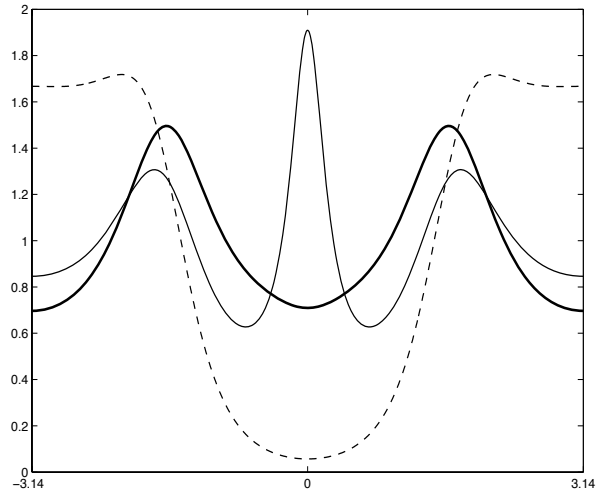


Fig. 1. Graphics of $\Psi(e^{i\omega})$ (dashed line), $\Phi^o(e^{i\omega})$ (bold line), and $\Phi^o_{KL}(e^{i\omega})$ (solid line) as functions of ω.

Simulation 2. Consider the following instance of Example 2 (corresponding to a Nevanlinna-Pick interpolation

problem):

$$A := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

and

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4/3 & 6/5 \\ 1 & 6/5 & 9/8 \end{bmatrix}$$

After the renormalization described at the beginning of Section 4 to transform Σ into the identity, we get

$$A' = \begin{bmatrix} -0.4378 & -0.2433 & -0.4370 \\ -0.0592 & 0.8659 & 0.2793 \\ 0.6724 & -0.3326 & 0.4052 \end{bmatrix}, \quad B' = \begin{bmatrix} 0.7471 \\ 0.4106 \\ 0.5226 \end{bmatrix}$$

Let $\Psi(z) = W_\psi(z)W_\psi^*(z)$, where

$$W_\psi(z) := K[1\ -1] \left(zI - \begin{bmatrix} -0.9 & 0 \\ 0 & 0.8 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

and $K = 0.3813$ so that $\int \Psi = 1$.

The algorithm converged to the matrix

$$\hat{\Lambda}' = \begin{bmatrix} 0.0650 & 0.1033 & -0.0511 \\ 0.1033 & 0.1097 & -0.0507 \\ -0.0511 & -0.0507 & -0.2636 \end{bmatrix}$$

The optimal solution $\Phi^o(e^{i\omega})$ is depicted (bold line) together with $\Psi(e^{i\omega})$ (dashed line) and the optimal constrained Kullback-Leibler approximation of Ψ (solid line) in Figure 2.

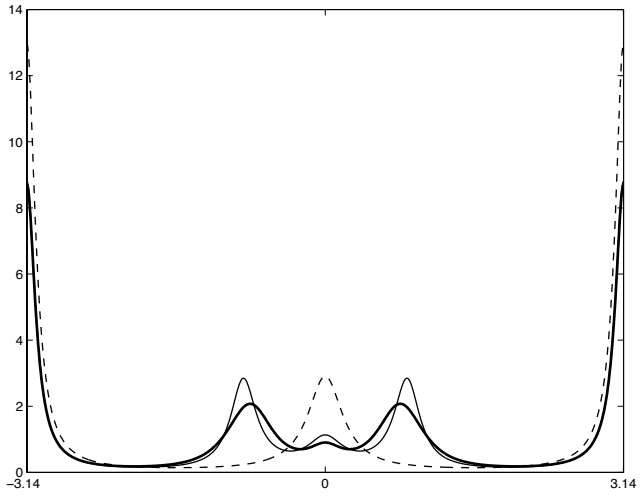


Fig. 2. Graphics of $\Psi(e^{i\omega})$ (dashed line), $\Phi^o(e^{i\omega})$ (bold line), and $\Phi^o_{KL}(e^{i\omega})$ (solid line) as functions of ω.

Simulation 3. Consider now the same case as in the Simulation 2 but with a different reference spectral density Ψ. Namely, let $\Psi(z) = W_\psi(z)W_\psi^*(z)$, where

$$W_\psi(z) := [0\ K] \left(zI - \begin{bmatrix} 5/100 & 1/4 \\ -1 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Take $K = 0.9675$ to ensure $\int \Psi = 1$ (notice that W has complex poles).

The algorithm converged to the matrix

$$\hat{\Lambda}' = \begin{bmatrix} 0.0151 & -0.0114 & 0.0036 \\ -0.0114 & -0.0561 & 0.0105 \\ 0.0036 & 0.0105 & 0.0340 \end{bmatrix}$$

The optimal solution $\Phi^o(e^{i\omega})$ is depicted (bold line) together with $\Psi(e^{i\omega})$ (dashed line) and the optimal constrained Kullback-Leibler approximation of Ψ (solid line) in Figure 3.

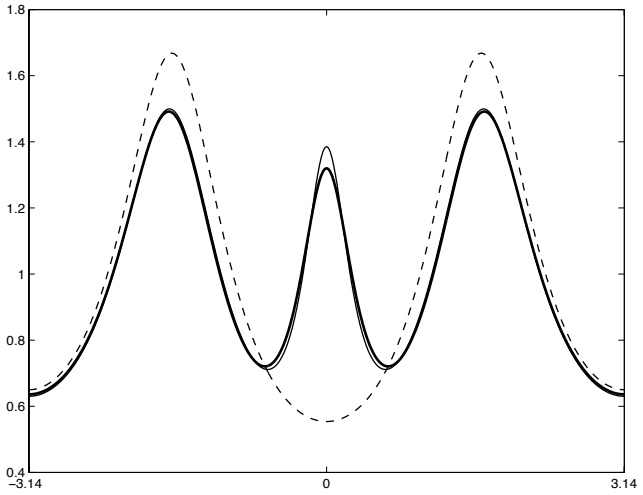


Fig. 3. Graphics of $\Psi(e^{i\omega})$ and $\Phi^o(e^{i\omega})$ (bold line) as functions of ω.

It will be interesting to compare the two approaches as spectral estimation methods, in the sense of [9].

VI. ACKNOWLEDGMENT

The authors wish to acknowledge their correspondence with prof. T. Georgiou that led to a considerable improvement of their understanding of distances between spectral density functions.

REFERENCES

[1] T. Georgiou and A. Lindquist, Kullback-Leibler approximation of spectral density functions, *IEEE Trans. Inform. Theory*, vol.49, pp. 2910-2917, 2003.

[2] T. Georgiou and A. Lindquist, Remarks on control design with degree constraint, *IEEE Trans. Aut. Control*, to be published. Available on line at http://www2.math.kth.se/ alq/GL2.pdf

[3] T. Georgiou, Solution of the general moment problem via a one-parameter imbedding,*IEEE Trans. Aut. Control*, Vol. 50, (2005), 811-826.

[4] T. Georgiou, Relative entropy and the multivariable multidimensional moment problem, *IEEE Trans. Inform. Theory*, vol.52, pp. 1052-1066, 2006.

[5] T. Georgiou, Distances between power spectral densities, arXiv e-print `math.OC/0607026`.

[6] T. Georgiou, An intrinsic metric for power spectral density functions, arXiv e-print `math.OC/0608486`.

[7] M. Pavon and A. Ferrante, On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities, *IEEE Trans. Aut. Control*, **51**, (2006), 639- 644.

[8] C. I. Byrnes, T. Georgiou, and A. Lindquist, A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint, *IEEE Trans. Automatic Control* vol. 46, pp. 822-839, 2001.
[9] C. I. Byrnes, T. Georgiou, and A. Lindquist, A new approach to spectral estimation: A tunable high-resolution spectral estimator, *IEEE Trans. Signal Processing* vol. 49, pp. 3189-3205, 2000.
[10] C. I. Byrnes, S. Gusev, and A. Lindquist, A convex optimization approach to the rational covariance extension problem, *SIAM J. Control and Opimization*, vol. 37, pp. 211-229, 1999.
[11] C. I. Byrnes, S. Gusev, and A. Lindquist, From finite covariance windows to modeling filters: A convex optimization approach, *SIAM Review*, vol. 43, pp. 645-675, 2001.
[12] C. I. Byrnes and A. Lindquist, *The generalized moment problem with complexity constraint*, Integral Equations and Operator Theory vol. **56(2)**, pp.163-180, 2006 (published online March 29, 2006).
[13] A. Ferrante, M. Pavon and F. Ramponi, *Hellinger vs. Kullback-Leibler multivariable spectrum approximation*, submitted for publication.
[14] A. Ferrante, M. Pavon and F. Ramponi, *Further results on the Byrnes-Georgiou-Lindquist generalized moment problem*, preprint March 2007, to appear.
[15] H. Cramér, *Mathematical methods of statistics* , Princeton Univ. Press (1946).
[16] H. Kuo, *Gaussian measures on Banach spaces* , Springer (1975).
[17] I.A. Ibragimov, R.Z. Has'minskii, *Statistical estimation: asymptotic theory*, Springer (1981) (Translated from Russian).
[18] L. Le Cam, Asymptotic methods in statistical decision theory, Springer Lectures Notes in Mathematics n. 976, p. 76-265, 1986.
[19] T. Georgiou, A topological approach to Nevanlinna-Pick interpolation, *SIAM J. on Mathematical Analysis*, vol. 18, pp. 1248-1260, 1987.
[20] T. Georgiou, The interpolation problem with a degree constraint, *IEEE Trans. on Automatic Control*, vol. 44, pp. 631-635, 1999.
[21] T. Georgiou, Spectral estimation by selective harmonic amplification, *IEEE Trans. Automat. Contr.*, vol.46, pp.29-42, 2001.
[22] T. Georgiou, The structure of state covariances and its relation to the power spectrum of the input, *IEEE Trans. Automat. Contr.*, vol.47, pp. 1056-1066, 2002.
[23] P. Enquist, A homotopy approach to rational covariance extension with degree constraint, *Int. J. Appl. Math. and Comp. Sci.*, vol. 11, pp.1173-1201, 2001.
[24] R. Nagamune, A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint, *Ieee Trans. Aut. Contr.*, vol 48, pp. 113-117, 2003.
[25] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004