Kullback–Leibler Approximation of Spectral Density Functions

Tryphon T. Georgiou, *Fellow, IEEE,* and Anders Lindquist, *Fellow, IEEE*

***Abstract*—We introduce a Kullback–Leibler-type distance between spectral density functions of stationary stochastic processes and solve the problem of optimal approximation of a given spectral density Ψ by one that is consistent with prescribed second-order statistics. In general, such statistics are expressed as the state covariance of a linear filter driven by a stochastic process whose spectral density is sought. In this context, we show i) that there is a unique spectral density Φ which minimizes this Kullback–Leibler distance, ii) that this optimal approximate is of the form Ψ/Q where the "correction term" Q is a rational spectral density function, and iii) that the coefficients of Q can be obtained numerically by solving a suitable convex optimization problem. In the special case where $\Psi = 1$, the convex functional becomes quadratic and the solution is then specified by linear equations.**

***Index Terms*—Approximation of power spectra, cross-entropy minimization, Kullback–Leibler distance, mutual information, optimization, spectral estimation.**

I. Introduction

In this paper, we are interested in approximation of power spectra of stationary stochastic processes and, in particular, in the following type of problem. Suppose that an *a priori* estimate Ψ of the power spectrum is available and that new data is obtained that is inconsistent with this estimate. Then the basic problem is to find another power spectrum Φ that is as close as possible to Ψ in some suitable sense and is consistent with the data. This is motivated by applications where one is called to reconcile inconsistent sets of data and incorporate prior information into the modeling process.

Thus, the starting point is a stationary stochastic process $\{y(t), t \in \mathbb{Z}\}$ with zero mean and spectral density Φ. In this paper, the spectral density will be regarded as a real-valued function $e^{i\theta} \mapsto \Phi(e^{i\theta})$ on the unit circle $\mathbb{T} := \{e^{i\theta} \mid -\pi \leq \theta < \pi\}$. The data for the approximation problem will be in the form of second-order statistics of $y(t)$. These could be estimates of autocorrelation lags $c_k := \mathrm{E}\{y(t)y(t+k)\}$ or, more often, cross correlations of outputs of different linear filters driven by $y(t)$. Such filters could represent the dynamics of physical or algorithmic devices used for measurement. A general framework encompassing all such possibilities is to take as data an estimate of the state covariance Σ of a linear system driven by $y(t)$. Thus, we require Φ to be consistent with such a state covariance.

As a distance measure between two spectral density functions, Ψ and Φ, we will use

$$\mathbb{S}(\Psi\|\Phi) := \int \Psi \log\left(\frac{\Psi}{\Phi}\right) \tag{1.1}$$

where, for economy of notation, we often write simply $\int \Omega$ to denote integrals of the form $\int_{-\pi}^{\pi} \Omega(e^{i\theta})\frac{d\theta}{2\pi}$. This is known as the *Kullback–Leibler distance*, originally applied to probability distributions [11], and possessing some rather useful properties. Although not symmetric in its arguments, it is jointly convex. Assuming that Ψ and Φ have the same zeroth moment, $\mathbb{S}(\Psi\|\Phi) \geq 0$ with equality if and only if $\Phi = \Psi$.

Given an arbitrary *a priori* spectral estimate Ψ and the second-order statistics Σ, the main result of this paper is that the problem of minimizing $\mathbb{S}(\Psi\|\Phi)$ subject to consistency with Σ has a unique minimum Φ. It is worth noting that this minimizing solution is unaffected by scaling Ψ, since this modifies $\mathbb{S}(\Psi\|\Phi)$ by a constant additive term and a positive, multiplicative, and constant factor.

When there is no *a priori* spectral estimate, it is reasonable to take $\Psi = 1$, which corresponds to white noise. Then, since

$$\mathbb{S}(1\|\Phi) := -\int \log \Phi$$

minimizing the Kullback–Leibler distance amounts to maximizing the entropy gain. Consequently, the maximum-entropy solution is the spectral density Φ closest to white noise that is consistent with the data.

The Kullback–Leibler integral has been studied intensively in statistics, information theory, and communication, typically optimizing with respect to the first argument Ψ [11]–[13], and has been considered as providing a distance measure between Gauss–Markov [29] and hidden Markov models [15]. In another direction, estimating Markov models from covariance data has a long history; see, e.g., [2], [8], [14], [17], [28], [30], [32], [36]. The approach taken here is based on ideas developed in [5], [6] in the context of Carathéodory extension, in [3], [4] in the context of Nevanlinna–Pick interpolation, and in [9], [10] in the context of generalized moment problems with complexity constraint. In this paper, the theory is developed in the framework of generalized analytic interpolation [21], [22].

II. State Covariance Statistics

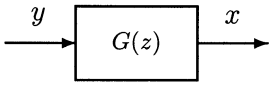
In this section, we explain the nature of the covariance data of our approximation problem. We model the measuring device as a linear system

Manuscript received May 2, 2002; revised January 5, 2003. This work was supported in part by grants from AFOSR, the Swedish Research Council, and the Göran Gustafsson Foundation.

T. T. Georgiou is with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455, USA (e-mail: tryphon@ece.umn.edu).

A. Lindquist is with the Department of Mathematics, Division of Optimization and Systems Theory, Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: alq@math.kth.se).

Communicated by A. Kavčić, Associate Editor for Detection and Estimation.
Digital Object Identifier 10.1109/TIT.2003.819324



with transfer function

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

such that A has all its eigenvalues in the open unit disc and (A, B) is a reachable pair, i.e.,

$$\operatorname{rank}[B, AB, A^2B, \dots, A^{n-1}B] = n.$$

Now, a general framework for obtaining second-order data is to pass the signal $y(t)$ through the device until it is in steady state and estimate the state covariance

$$\Sigma := \mathrm{E}\{x(t)x(t)^*\}.$$

In fact, if the input is a stationary stochastic process $\{y(t), t \in \mathbb{Z}\}$, the output process $\{x(t), t \in \mathbb{Z}\}$ is a stationary vector process satisfying the state equation

$$x(t) = Ax(t-1) + By(t). \tag{2.1}$$

Therefore, in order for the spectral density Φ of $\{y(t),\ t \in \mathbb{Z}\}$ to be consistent with the data, it must satisfy the constraint

$$\int G\Phi G^* = \Sigma. \tag{2.2}$$

Example 1: The simplest case is when the data consists of a finite sequence of covariance lags

$$c_k = \mathrm{E}\{y(t)\overline{y}(t+k)\}, \qquad k = 0, 1, \dots, n-1.$$

This corresponds to

$$G(z) = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-2} \\ z^{n-1} \end{bmatrix} \quad \text{with } A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

and the state covariance

$$\Sigma = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ \bar{c}_1 & c_0 & \cdots & c_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{c}_{n-1} & \bar{c}_{n-2} & \cdots & c_0 \end{bmatrix} \tag{2.3}$$

is a Toeplitz matrix.

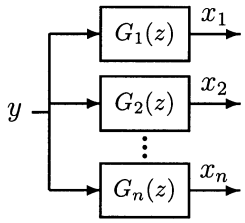
The problem of parameterizing the class of spectral densities that satisfy (2.2) for this case is the classical trigonometric moment problem going back to Carathéodory, Schur, and Toeplitz (see, e.g., [25]). This class is nonempty if and only if $\Sigma \geq 0$ (which here of course is always the case since Σ is assumed to be a covariance), and it contains infinitely many elements if and only if $\Sigma > 0$. Such spectral densities correspond to infinite extensions of the finite covariance sequence $c_0, c_1, \dots, c_{n-1}$, and hence the problem is also referred to as the *covariance extension problem*.

In particular, the so-called *maximum-entropy extension* (see, e.g., [26]) is the unique covariance extension that maximizes the entropy gain

$$\int \log \Phi = -\mathbb{S}(1\|\Phi)$$

and is obtained in our framework as the minimizer of (1.1) when $\Psi = 1$. When Ψ ranges over *arbitrary* positive trigonometric polynomials of degree n, the minimizers of (1.1) are precisely the coercive spectral densities of degree at most $2n$ that are consistent with the data, thus recovering basic results in [5]–[8], [18], [20].

More generally, the measurement device could be a bank of filters



with transfer function

$$G(z) = \begin{bmatrix} G_1(z) \\ G_2(z) \\ \vdots \\ G_n(z) \end{bmatrix}. \tag{2.4}$$

The following are two examples of particular interest that have been explored in the context of high-resolution spectral estimation in [3], [4], [21].

Example 2: First, consider the case where

$$G_k(z) = \frac{1}{1 - p_k z}, \qquad k = 1, 2, \dots, n$$

with $p_1, p_2, \dots, p_n$ distinct. Then

$$A = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_n \end{bmatrix}, \qquad B = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

and the state covariance Σ can be shown to have the form of a Pick matrix

$$\sigma = \begin{bmatrix} \frac{w_1+\bar{w}_1}{1-p_1\bar{p}_1} & \frac{w_1+\bar{w}_2}{1-p_1\bar{p}_2} & \cdots & \frac{w_1+\bar{w}_n}{1-p_1\bar{p}_n} \\ \frac{w_2+\bar{w}_1}{1-p_2\bar{p}_1} & \frac{w_2+\bar{w}_2}{1-p_2\bar{p}_2} & \cdots & \frac{w_2+\bar{w}_n}{1-p_2\bar{p}_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n+\bar{w}_1}{1-p_n\bar{p}_1} & \frac{w_n+\bar{w}_2}{1-p_n\bar{p}_2} & \cdots & \frac{w_n+\bar{w}_n}{1-p_n\bar{p}_n} \end{bmatrix} \tag{2.5}$$

with

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

(cf. Remark 7 later).

The problem of parameterizing the class of spectral densities that satisfy (2.2) for this case now relates to the classical Nevanlinna–Pick interpolation problem (see, e.g., [1], [40]). In this

problem, one seeks an analytic function F on the unit disc $\mathbb{D}$ that has positive real part and satisfies the interpolation conditions $F(\bar{p}_k) = \bar{w}_k$, $k = 1, 2, \dots, n$. The data $w_1, w_2, \dots, w_n$ can be recovered (modulo a common imaginary constant that plays no role) from the measured covariance statistics Σ, and the spectral density Φ can be obtained from F via $\Phi(e^{i\theta}) = 2\mathrm{Re}\{F(e^{i\theta})\}$.

Following classical theory, the class of solutions is nonempty if and only if $\Sigma \geq 0$ (which again is guaranteed if Σ is to be a covariance), and it contains infinitely many elements if and only if $\Sigma > 0$. The minimizer of (1.1) when $\Psi = 1$ is the so-called *central solution*. Moreover, suitable choices of Ψ as a rational function having the same poles as the filter bank, leads to a complete parameterization of all rational spectral densities of degree at most $2n$ that are consistent with the data, thus recovering results from [3], [4], [19], [20].

Example 3: In the case where

$$G_k(z) = \frac{1}{(1-pz)^{k-1}}, \qquad k = 1, 2, \dots, n$$

the system matrices become

$$A = \begin{bmatrix} p & 1 & 0 & \cdots & 0 \\ 0 & p & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & p \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

while the covariance matrix is now of the form

$$\Sigma = \frac{1}{2}(WE + EW^*)$$

where W is an upper triangular Toeplitz matrix, and E is the reachability gramian of (A, B) [21, p. 33]. In the same way as above, spectral densities consistent with the data do exist if and only if $\Sigma \geq 0$, and there are infinitely many when $\Sigma > 0$. Analogous results and parameterization of spectral densities which are consistent with covariance data can be obtained in this case as well.

The family of spectral densities consistent with the state statistics Σ in this example, as well as in Example 2, strongly depends on the choice of poles $p_1, p_2, \dots, p_n$. In fact, by choosing these filter bank poles appropriately, it is possible to give preference to selected frequency bands of the spectrum and allow more accurate reconstruction of the spectral density in these particular frequency bands [3], [4], [21], [34].

These examples suggest that, besides being Hermitian, Σ is highly structured and that there is a redundancy in the formulation of the constraint (2.2). In fact, Σ is in the range of an integral operator

$$\Gamma : \Phi \mapsto \Sigma = \int G\Phi G^* \tag{2.6}$$

which, in our present setting, it suffices to define on the space $C(\mathbb{T})$ of continuous functions on the unit circle $\mathbb{T}$. The range of Γ can then be described as follows [21], [22]. A matrix Σ is in the range of Γ if and only if

$$\Sigma - A\Sigma A^* = BH + H^*B^* \tag{2.7}$$

has a solution $H \in \mathbb{C}^{1\times n}$, or, equivalently

$$\mathrm{rank}\begin{pmatrix} \Sigma - A\Sigma A^* & B \\ B^* & 0 \end{pmatrix} = \mathrm{rank}\begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}. \tag{2.8}$$

Moreover, any positive semidefinite matrix Σ which satisfies (2.7) qualifies as a state covariance of the system (2.1) for a suitable stationary input process [22, Theorem 1]. In Section VII, we shall need the following lemma.

Lemma 4: The range of the operator $\Gamma : C(\mathbb{T}) \to \mathbb{C}^{n\times n}$, defined by (2.6), has real dimension $2n - 1$.

Proof: Since the linear map $\Sigma \mapsto \Sigma - A\Sigma A^*$ is invertible, (2.7) has a solution for all $H \in \mathbb{C}^{1\times n}$, and this solution is zero if and only if $BH + H^*B^* = 0$. Therefore, it only remains to prove that the kernel $\{H \in \mathbb{C}^{1\times n} \mid BH + H^*B^* = 0\}$ has real dimension one.

To this end, first note that any $H \in \mathbb{C}^{1\times n}$ can be written as $H = (\beta + i\alpha)B^* + w$, where $\alpha, \beta \in \mathbb{R}$ and $wB = 0$, and hence $BH + H^*B^* = 0$ if and only if

$$\Omega := 2\beta BB^* + Bw + w^*B^* = 0.$$

Now, it follows from $B^*\Omega B = 0$ that $\beta = 0$, which inserted into $\Omega B = 0$ yields $w = 0$. Consequently, $BH + H^*B^* = 0$ if and only if $H = i\alpha B^*$ for some $\alpha \in \mathbb{R}$, proving that the kernel has real dimension one, as claimed. □

III. On the Kullback–Leibler Measure

The Kullback–Leibler distance has its roots in hypothesis testing and represents the mean information per observation for discrimination of a possible underlying probability density function from another [31, p. 6]. It is also called cross-entropy [24] or gain of information [37]. In this paper, we use $\mathbb{S}(\Psi\|\Phi)$ as measure of distance between spectral density functions. To this end, several remarks are in order.

If $\int \Psi$ equals $\lambda := \int \Phi$, then, by Jensen's inequality

$$\mathbb{S}(\Psi\|\Phi) = \lambda\mathbb{S}(\frac{\Psi}{\lambda}\|\frac{\Phi}{\lambda}) \geq 0$$

with equality if and only if $\Phi = \Psi$. Although $\mathbb{S}(\Psi\|\Phi)$ is not a metric, this property allows us to interpret it as a distance between Φ and Ψ.

It is often the case that λ is part of the data Σ. This happens exactly when A is a singular matrix. To see this, let $v^*A = 0$ for some $v \in \mathbb{C}^n$. Then

$$v^*\Sigma v = \int v^*G\Phi G^*v = \int v^*B\Phi B^*v = \|v^*B\|^2 \int \Phi.$$

Thus, $\lambda = v^*\Sigma v/\|v^*B\|^2$.

Therefore, in case A is singular, Ψ can be scaled so that $\int \Psi = \lambda$. Then $\mathbb{S}(\Psi\|\Phi)$ is a *bona fide* distance measure, and it is reasonable to consider the problem of minimizing $\mathbb{S}(\Psi\|\Phi)$ over Φ subject to the constraints $\Gamma(\Phi) = \Sigma$.

However, even in the case where A is nonsingular, the minimization problem makes sense, and, as we shall see in the next section, it has a unique solution $\hat{\Phi}$. If we normalize Ψ, *a posteriori*, to the new function $\hat{\Psi} = (\int \hat{\Phi}/\int \Psi)\Psi$, then

$$\mathbb{S}(\hat{\Psi}\|\Phi) = \alpha\mathbb{S}(\Psi\|\Phi) + \beta$$

where $\alpha = \int \hat{\Phi} / \int \Psi$ and $\beta = \int \hat{\Phi} \log \alpha$. Thus, the minimizing Φ would be the same in either case, and $\mathbb{S}(\hat{\Psi}||\Phi)$ would retain a distance-like character, being nonnegative except when the two arguments are identical in which case it is zero.

IV. The Approximation Problem

Denote by $C_+(\mathbb{T})$ the subset of positive functions in the space $C(\mathbb{T})$ of continuous functions on the unit circle $\mathbb{T}$. We now state and prove our main result.

Theorem 5: Let (A, B) be a reachable pair with the eigenvalues of A in the open unit disc, and let $G(z) = (I - zA)^{-1}B$. Moreover, let Γ be defined by (2.6), and let

$$\mathfrak{L}_+ := \{\Lambda \in \text{range}\,(\Gamma) \mid G(e^{i\theta})^* \Lambda G(e^{i\theta}) > 0, \theta \in [-\pi, \pi]\}.$$

Given any $\Psi \in C_+(\mathbb{T})$ and any $\Sigma > 0$ which satisfies (2.8), there is a unique $\Phi \in C_+(\mathbb{T})$ that minimizes

$$\mathbb{S}(\Psi||\Phi) = \int_{-\pi}^{\pi} \Psi(e^{i\theta}) \log \frac{\Psi(e^{i\theta})}{\Phi(e^{i\theta})} \frac{d\theta}{2\pi} \tag{4.1}$$

subject to $\Gamma(\Phi) = \Sigma$, i.e., subject to

$$\int_{-\pi}^{\pi} G(e^{i\theta}) \Phi(e^{i\theta}) G(e^{i\theta})^* \frac{d\theta}{2\pi} = \Sigma. \tag{4.2}$$

This minimizer takes the form

$$\hat{\Phi} = \frac{\Psi}{G^* \hat{\Lambda} G} \tag{4.3}$$

where $\hat{\Lambda}$ is the unique solution to the problem of minimizing

$$\mathcal{J}_\Psi(\Lambda) := \text{tr}(\Lambda\Sigma) - \int \Psi \log G^* \Lambda G \tag{4.4}$$

over all $\Lambda \in \mathfrak{L}_+$. Here $\text{tr}\, M$ denotes the trace of the matrix M.

The techniques used in proving this theorem are based on duality theory of mathematical programming and are motivated by the ideas in [5], [6], [3], [4], [9].

We begin by forming the Lagrangian

$$L(\Phi, \Lambda) = \mathbb{S}(\Psi||\Phi) + \text{tr}\Big(\Lambda\big(\Gamma(\Phi) - \Sigma\big)\Big).$$

Here $\Lambda \in \text{range}\,(\Gamma)$ is the Lagrange multiplier. Next, we consider the problem of maximizing the dual functional

$$\Lambda \mapsto \inf_{\Phi \in C_+(\mathbb{T})} L(\Phi, \Lambda).$$

Note that

$$\begin{aligned} L(\Phi, \Lambda) &= \int \Psi \log \frac{\Psi}{\Phi} + \text{tr}\left(\Lambda\left[\int G\Phi G^* - \Sigma\right]\right) \\ &= \int \Psi \log \frac{\Psi}{\Phi} + \int G^* \Lambda G \Phi - \text{tr}(\Lambda\Sigma). \end{aligned}$$

The dual functional will take finite values only if

$$Q(e^{i\theta}) := G(e^{i\theta})^* \Lambda G(e^{i\theta}) > 0, \qquad \theta \in [-\pi, \pi].$$

Therefore, without loss of generality, we restrict Λ to the set

$$\mathfrak{L}_+ := \{\Lambda \in \text{range}\,(\Gamma) \mid Q(e^{i\theta}) > 0, \theta \in [-\pi, \pi]\}.$$

For each $\Lambda \in \mathfrak{L}_+$, the function $\Phi \mapsto L(\Phi, \Lambda)$ is strictly convex. Hence, if this function has a minimum in $C_+(\mathbb{T})$, then the directional derivative

$$\delta L(\Phi, \Lambda; \delta\Phi) = \int \delta\Phi \left(Q - \frac{\Psi}{\Phi}\right)$$

will be zero for all continuous variations $\delta\Phi$ at the minimum. Therefore, the possible minimizer must be of the form

$$\Phi = \frac{\Psi}{Q}$$

and the dual functional must be

$$L(\frac{\Psi}{Q}, \Lambda) = -\mathcal{J}_\Psi(\Lambda) + \int \Psi$$

with $\mathcal{J}_\Psi(\Lambda)$ defined by (4.4). Hence, the dual optimization problem amounts to minimizing $\mathcal{J}_\Psi(\Lambda)$ over $\mathfrak{L}_+$. The following key lemma will be proved in the next section.

Lemma 6: The functional $\mathcal{J}_\Psi(\Lambda)$ has a unique minimum $\hat{\Lambda} \in \mathfrak{L}_+$. Moreover

$$\Gamma(\frac{\Psi}{G^* \hat{\Lambda} G}) = \Sigma.$$

We now use this lemma to complete the proof of the theorem. We first note that the function in the lemma, given by

$$\hat{\Phi} = \frac{\Psi}{\hat{Q}}, \qquad \hat{Q} = G^* \hat{\Lambda} G$$

belongs to $C_+(\mathbb{T})$ and is a stationary point of $\Phi \mapsto L(\Phi, \hat{\Lambda})$, which is strictly convex. Consequently

$$L(\hat{\Phi}, \hat{\Lambda}) \le L(\Phi, \hat{\Lambda}), \qquad \text{for all } \Phi \in C_+(\,\mathbb{T}\,)$$

or, equivalently, since $\Gamma(\hat{\Phi}) = \Sigma$

$$\mathbb{S}(\Psi||\hat{\Phi}) \le \mathbb{S}(\Psi||\Phi)$$

for all $\Phi \in C_+(\mathbb{T})$ satisfying the constraint $\Gamma(\Phi) = \Sigma$. The above holds with equality if and only if $\Phi = \hat{\Phi}$. This completes the proof of the theorem.

Remark 7: The connection between the approximation problem of this section to analytic interpolation problems described in [3], [4] is in terms of the function

$$F(z) = \frac{1}{4\pi} \int \frac{e^{i\theta} + z}{e^{i\theta} - z} \Phi(e^{i\theta}) d\theta$$

which is analytic in $\mathbb{D}$ and takes values in the open left half of the complex plane. This so-called Carathéodory function can be continuously extended to the unit circle $\mathbb{T}$, where Φ is continuous, and

$$\Phi = 2\text{Re}\{F\} = F + F^*$$

there. This is the content of the Riesz–Herglotz theorem [27]. Now, since $G = B + zAG$

$$\int GFG^* = B\left(\int FG^*\right) + \int e^{i\theta} AGFB + A\left(\int GFG^*\right) A^*$$

where the middle term is zero since GF is analytic in $\mathbb{D}$. Therefore, $\Sigma := \int G\Phi G^*$ satisfies (2.7) if and only if

$$H = \int FG^*. \tag{4.5}$$

Consequently, the constraint (4.2) in terms of Φ is equivalent to the constraint (4.5) in terms of F [22].

V. The Dual Problem

To prove Lemma 6, we need to consider the dual problem to minimize $\mathcal{J}_\Psi(\Lambda)$ over $\mathfrak{L}_+$. To this end, form the directional derivative

$$\delta\mathcal{J}_\Psi(\Lambda;\delta\Lambda) = \operatorname{tr}\left(\delta\Lambda\left[\Sigma - \int G\frac{\Psi}{Q}G^*\right]\right) \tag{5.1}$$

in any direction $\delta\Lambda \in \operatorname{range}(\Gamma)$, as well as the second derivative

$$\delta^2\mathcal{J}_\Psi(\Lambda;\delta\Lambda) = \int \frac{\Psi}{Q^2}(G^*\delta\Lambda G)^2. \tag{5.2}$$

From these expressions we observe two things. First, any stationary point of $\mathcal{J}_\Psi(\Lambda)$ must satisfy the equation

$$\omega(\Lambda) = \Sigma \tag{5.3}$$

where the map $\omega : \mathfrak{L}_+ \to \mathfrak{S}_+$ between $\mathfrak{L}_+$ and $\mathfrak{S}_+ := \{\Sigma \in \operatorname{range}(\Gamma) \mid \Sigma > 0\}$ is defined as

$$\Lambda \mapsto \int G\frac{\Psi}{Q}G^*.$$

Second, $\mathcal{J}_\Psi(\Lambda)$ is strictly convex, and therefore there is at most one such stationary point, which would then be the unique minimum. This is an immediate consequence of (5.2) and the following lemma. In particular, the map $\omega : \mathfrak{L}_+ \to \mathfrak{S}_+$ is injective.

Lemma 8: Suppose $\Lambda \in \operatorname{range}(\Gamma)$ is such that $G^*\Lambda G = 0$ on $\mathbb{T}$. Then $\Lambda = 0$.

Proof: Without loss of generality, we assume that G is normalized so that $\int GG^* = I$. This can be achieved by a similarity transformation of the system matrices (A, B). Suppose $G^*\Lambda G = 0$. Then

$$\int e^{-ik\theta}G^*\Lambda G = 0, \quad \text{for all } k.$$

Since $\Lambda \in \operatorname{range}(\Gamma)$, and therefore $\Lambda = \int G\Theta G^*$ for some $\Theta \in C(\mathbb{T})$, this can be written

$$\int e^{-ik\theta}G(e^{i\theta})^*\left(\int G(e^{i\varphi})\Theta(e^{i\varphi})G(e^{i\varphi})^*d\varphi\right)G(e^{i\theta})d\theta = 0.$$

By changing the order of integration and using the fact that $\int e^{-ik\theta}GG^* = A^k$ for $k \geq 0$, we deduce that

$$0 = \int G^*A^kG\Theta = \int e^{-ik\theta}G^*G\Theta = \int e^{-ik\theta}\|G\|^2\Theta.$$

A similar argument, exchanging A for A^*, shows that

$$\int e^{-ik\theta}\|G\|^2\Theta = 0$$

also for negative k. Thus, $\|G\|^2\Theta = 0$ on $\mathbb{T}$. By reachability of (A, B), $\|G\| \neq 0$ on $\mathbb{T}$, and hence, $\Theta = 0$, and so is Λ. □

Next, we prove that in fact $\omega : \mathfrak{L}_+ \to \mathfrak{S}_+$ is also surjective, thus establishing the existence of a unique solution to (5.3), and hence a unique minimum of the dual functional $\mathcal{J}_\Psi$. To this end, we first note that both sets $\mathfrak{L}_+$ and $\mathfrak{S}_+$ are nonempty, convex, and open subsets of the same Euclidean space, and hence diffeomorphic to this space. Besides being continuous, the map ω is also proper, i.e., the inverse image $\omega^{-1}(K)$ is compact for any compact K in $\mathfrak{S}_+$. To see this, first note that each $\omega^{-1}(K)$ must be bounded. In fact, $\det\omega(\Lambda) \to 0$ as $\|\Lambda\| \to \infty$, and similarly $\det\omega(\Lambda) \to \infty$ as Λ tends to the boundary of $\mathfrak{L}_+$. Since, therefore, $\omega : \mathfrak{L}_+ \to \mathfrak{S}_+$ is a proper and injective continuous map between connected spaces of the same dimension, it follows from [9, Lemma 2.3] that ω is a homeomorphism. This completes the proof of Lemma 6.

Consequently, the dual problem provides us with an approach to computing the unique Φ that minimizes the Kullback–Leibler distance $\mathbb{S}(\Psi\|\Phi)$ subject to the constraint $\Gamma(\Phi) = \Sigma$.

VI. The Maximum Entropy Solution

We now turn to the case where $\Psi = 1$. As we explained before, this represents the situation when no *a priori* spectral estimate is available, and the selected spectral density Ψ corresponds to white noise. The Kullback–Leibler distance

$$\mathbb{S}(1\|\Phi) := -\int \log\Phi$$

now has the interpretation of entropy rate of the underlying process [26].

In this case, the solution of the dual problem is particularly simple, and we shall derive a closed-form solution along similar lines to those in [16], [33]. In fact, the dual functional becomes

$$\mathcal{J}_1(\Lambda) = \operatorname{tr}(\Lambda\Sigma) - \int \log G^*\Lambda G \tag{6.1}$$

where the first term can be rewritten as

$$\begin{aligned}\operatorname{tr}(\Lambda\Sigma) &= \operatorname{tr}(\Lambda\int G\Phi G^*)\\ &= \int \operatorname{tr}(\Lambda G\Phi G^*)\\ &= \int G^*\Lambda G\Phi.\end{aligned}$$

To simplify the stationarity conditions of $\mathcal{J}_1$ we use an alternative representation for

$$Q := G^*\Lambda G$$

which is the content of the following lemma.

Lemma 9: If $Q(e^{i\theta}) > 0$ for $\theta \in [-\pi, \pi]$, then there is a $C \in \mathbb{C}^{n\times 1}$ such that C^*G is an outer function and

$$Q = C^*GG^*C.$$

Proof: We first introduce notation. Let

$$\varphi(z) = \frac{\det(zI - A^*)}{\det(I - zA)}$$

and define

$$\mathcal{K} = H^2(\mathbb{T}) \ominus \varphi H^2(\mathbb{T})$$

where $H^2(\mathbb{T})$ denotes the Hardy space of square-integrable functions on the unit circle $\mathbb{T}$ with analytic continuation in the open unit disc $\mathbb{D}$. We remark that φ is an inner function and that $\mathcal{K}$ is finite-dimensional with the elements of G forming a basis. Hence, $\varphi Q \in H^1(\mathbb{T})$, where $H^1(\mathbb{T})$ is the space of integrable Hardy functions.

Since $Q = Q^* > 0$ on $\mathbb{T}$, there is an outer function $a \in H^2(\mathbb{T})$ such that $Q = a^*a$. Moreover, since $|\varphi| = 1$, $|\varphi Q| = |a|^2$. Therefore, the outer part of φQ equals a^2. Since the elements of $\varphi^* G$ are orthogonal to $H^2(\mathbb{T})$, it follows that φG^* vanishes at the origin, and so does φQ. Hence, $\varphi Q = \theta a^2$, where θ is inner and $\theta(0) = 0$. In view of the fact that $\varphi a^* = \theta a$, it readily follows that $\varphi^* a$ is orthogonal to $H^2(\mathbb{T})$. Therefore, $a \in \mathcal{K}$ and can thus be expressed as C^*G for some C. □

Although we are here interested in functions that are rational, the lemma actually holds in greater generality.

Returning to the dual functional, we now have

$$\begin{aligned}\mathcal{J}_1(\Lambda) &= \int C^*GG^*C\Phi - \int \log C^*GG^*C \\ &= \int C^*G\Phi G^*C - 2\int \log|C^*G| \\ &= C^*\Sigma C - 2\log C^*G(0)\end{aligned}$$

where in the last step we used the fact that C^*G is an outer function and thus that $\log C^*G$ is harmonic in $\mathbb{D}$ [38]. Since $G(0) = B$, the dual functional only depends on C, namely

$$J_1(C) := C^*\Sigma C - 2\log C^*B$$

which we now differentiate to obtain the equation

$$2\Sigma C - 2\frac{1}{C^*B}B = 0$$

for the minimizer. Therefore, $C = \frac{1}{B^*C}\Sigma^{-1}B$, and hence $(B^*C)^2 = B^*\Sigma^{-1}B$. Consequently, the maximum-entropy spectral density is given by

$$\Phi(e^{i\theta}) = \frac{B^*\Sigma^{-1}B}{B^*\Sigma^{-1}G(e^{i\theta})G(e^{i\theta})^*\Sigma^{-1}B}. \tag{6.2}$$

A version of this formula, valid for multivariable processes, appears in [23].

VII. On the Numerical Solution of the Dual Problem

The case when $\Psi = 1$ is very special, and, as we have seen, it can be solved in closed form. In general, when Ψ is an arbitrary rational, or even irrational, function, one needs an iterative algorithm for solving the dual problem. Since the dual functional $\mathcal{J}_\Psi$ is convex, such an algorithm could be based on Newton's method. For this, we first need a global coordinatization of $\mathfrak{L}_+$ that preserves convexity.

To this end, given a $\Lambda \in \mathfrak{L}_+$, let M be the unique solution of the Lyapunov equation

$$M = A^*MA + \Lambda. \tag{7.1}$$

Then, since

$$G^*\Lambda G = G^*MG - G^*A^*MAG$$

and $AG = z^{-1}(G - B)$, we have

$$Q := G^*\Lambda G = (G - \frac{1}{2}B)^*K + K^*(G - \frac{1}{2}B) \tag{7.2}$$

where $K := MB$. Let $\mathfrak{K}_+$ be the space of all $K \in \mathbb{C}^{n\times 1}$ such that $Q > 0$ on the unit circle $\mathbb{T}$. Since

$$B^*K = B^*MB = \int Q$$

must be real, $\mathfrak{K}_+$ has real dimension $2n - 1$.

Lemma 10: The linear map $\kappa : \mathfrak{L}_+ \to \mathfrak{K}_+$, defined by (7.1) and $K = MB$, is a bijection.

Proof: By Lemma 4 , $\mathfrak{L}_+$ has real dimension $2n - 1$, the same as that of $\mathfrak{K}_+$. Therefore, $\mathfrak{L}_+$ and $\mathfrak{K}_+$ are both nonempty open convex sets in $\mathbb{R}^{2n-1}$ and hence diffeomorphic to $\mathbb{R}^{2n-1}$. Since κ is a linear map between Euclidean spaces of the same dimension, it just remains to show that κ is injective. However, this follows immediately from (7.2) and Lemma 8. □

To express the dual functional $\mathcal{J}_\Psi$ in the new coordinates, we also observe that

$$\mathrm{tr}\{\Lambda\Sigma\} = \mathrm{tr}\{M\Sigma\} - \mathrm{tr}\{MA\Sigma A^*\}$$

by (7.1), which, in view of (2.7), becomes

$$\mathrm{tr}\{\Lambda\Sigma\} = HK + K^*H^*.$$

Consequently, the dual problem is equivalent to minimizing the convex functional

$$\hat{\mathcal{J}}_\Psi(K) := HK + K^*H^* - \int \Psi \log\Big((G - \frac{1}{2}B)^*K + K^*(G - \frac{1}{2}B)\Big) \tag{7.3}$$

over $\mathfrak{K}_+$. Differentiating $\hat{\mathcal{J}}_\Psi$ with respect to the complex variables $K_1, K_2, \ldots, K_n$, remembering that $\frac{\partial \bar{K}_j}{\partial K_\ell} = 0$ for all j, ℓ, we obtain the gradient

$$\frac{\partial \hat{\mathcal{J}}_\Psi}{\partial K} = H - \int \frac{\Psi}{Q}(G - \frac{1}{2}B)^* \tag{7.4}$$

with Q given by (7.2), which can also be written as

$$\frac{\partial \hat{\mathcal{J}}_\Psi}{\partial K} = H - \int FG^* \tag{7.5}$$

in terms of the Carathéodory function

$$F(z) = \frac{1}{4\pi}\int \left(\frac{e^{i\theta} + z}{e^{i\theta} - z}\right)\frac{\Psi(e^{i\theta})}{Q(e^{i\theta})}d\theta \tag{7.6}$$

(see Remark 7). To see this, observe that $\int F^*G^* = \int F^*B^* = \int FB^* = F(0)B^*$. We note that the gradient equals zero when the condition (4.5), or, equivalently, the condition (2.2), holds, in harmony with Theorem 5.

Now, the function $\hat{\mathcal{J}}_\Psi : \mathfrak{K}_+ \to \mathbb{R}$ is strictly convex, and therefore Newton's method can be used to determine its minimum. Such a method was developed in [4] for an analogous problem. The drawback seems to be that the gradient (7.4) becomes unbounded near the boundary of $\mathfrak{K}_+$, often causing numerical instability.

The alternative parameterization in terms of C, used in the previous section, is also applicable to the case of a general Ψ. To see this, replace $\mathrm{tr}(\Lambda\Sigma)$ by $C^*\Sigma C$ and $G^*\Lambda G$ by C^*GG^*C in (4.4). The dual functional is now expressed in terms of C in the form

$$J_\Psi(C) = C^*\Sigma C - \int \Psi \log(C^*GG^*C) \tag{7.7}$$

and the gradient

$$\frac{\partial J_\Psi}{\partial C} = C^*\Sigma - \int \frac{\Psi}{G^*C} G^* = C^*\left(\Sigma - \int G\frac{\Psi}{Q}G^*\right) \tag{7.8}$$

where $Q := C^*GG^*C$, is zero whenever condition (2.2) holds. This parameterization has two advantages. It avoids a time-consuming spectral factorization problem in each iteration, and also is better behaved at the boundary. However, J_Ψ is not globally convex. Such parameterizations have been explored in [16] for Carathéodory extension problems and in [33] for Nevanlinna–Pick interpolation. In these papers, a homotopy continuation method was used to compensate for the loss of global convexity.

VIII. Concluding Remarks

Given generalized covariance data Σ of a stationary stochastic process and an initial estimate Ψ for its spectral density, which may be inconsistent with the data, we formulated and solved the approximation problem of determining a closest approximant to Ψ in the sense of Kullback–Leibler which is also consistent with the data. In particular, we have shown that the minimizing function is unique.

This problem is relevant when statistics is specified in the form of a state covariance of a linear system driven by the unknown process. This is a rather general situation which, in particular, encompasses spectral analysis in linear arrays [39] with ordinary partial autocorrelation function, as well as spectral analysis using filter banks [3], [4].

The basic techniques that we have developed here should carry over to the case of a vector-valued stochastic processes, where the distance measure is replaced by the matricial Kullback–Leibler–von Neumann generalization [35]

$$\mathbb{S}(\Psi\|\Phi) := \mathrm{tr}\left(\int \Psi\log\Psi - \int \Psi\log\Phi\right).$$

In this case, the data can again be encapsulated as the state covariance of a multiple-input system [22]. This is the topic of a forthcoming paper.

References

[1] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*. New York: Hafner, 1965.

[2] A. Benveniste and C. Chaure, "AR and ARMA identification algorithms of levinson type: An identification approach," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 1243–1261, Dec. 1981.

[3] C. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for nevanlinna-pick interpolation: A convex optimization approach to certain problems in systems and control," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 822–839, June 2001.

[4] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Processing*, vol. 48, pp. 3189–3206, Nov. 2000.

[5] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM J. Control and Opt.*, vol. 37, pp. 211–229, 1999.

[6] ——, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM Rev.*, vol. 43, pp. 645–675, Dec. 2001.

[7] C. I. Byrnes, H. J. Landau, and A. Lindquist, "On the well-posedness of the rational covariance extension problem," in *Current and Future Directions in Applied Mathematics*, M. Alber, B. Hu, and J. Rosenthal, Eds. Basel, Switzerland: Birkhäser, 1997, pp. 83–106.

[8] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Mateev, "A complete parametrization of all positive rational extensions of a covariance sequence," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1841–1857, Nov. 1995.

[9] C. I. Byrnes and A. Lindquist, "Interior point solutions of variational problems and global inverse function theorems," manuscript submitted for publication, Tech. Rep. TRITA/MAT-01-OS13, 2001. Also [Online] at http://www.math.kth.se/alq/, submitted for publication.

[10] ——, "A convex optimization approach to generalized moment problems," in *Control and Modeling of Complex Systems: Cybernetics in the 21st Century*, K. Hashimoto, Y. Oishi, and Y. Yamamoto, Eds. Basel, Switzerland: Birkhäuser, 2003, pp. 3–21.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[12] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Probab.*, vol. 3, pp. 146–158, 1975.

[13] I. Csiszár and F. Matúš, "Information projections revisited," preprint, 2001.

[14] P. Dewilde and H. Dym, "Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 446–461, July 1981.

[15] M. N. Do, "Fast approximation of kullback-leibler distance for dependence trees and hidden markov models," Univ. Illinois at Urbana-Champaign, preprint.

[16] P. Enqvist, "A homotopy approach to rational covariance extension with degree constraint," *Int. J. Appl. Math. and Comp. Sci.*, vol. 11, no. 5, pp. 1173–1201, 2001.

[17] T. T. Georgiou, "Partial realization of covariance sequences," Ph.D. dissertation, CMST, Univ. Florida, Gainesville, 1983.

[18] ——, "Realization of power spectra from partial covariance sequences," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-35, pp. 438–449, Apr. 1987.

[19] ——, "A topological approach to nevanlinna-pick interpolation," *SIAM J. Math. Anal.*, vol. 18, pp. 1248–1260, 1987.

[20] ——, "The interpolation problem with a degree constraint," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 631–635, Mar. 1999.

[21] ——, "Spectral estimation by selective harmonic amplification," *IEEE Trans. Automat. Contr.*, vol. 46, pp. 29–42, Jan. 2001.

[22] ——, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 1056–1066, July 2002.

[23] ——, "Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parameterization," *IEEE Trans. Automat. Contr.*, vol. 47, pp. 1811–1823, Nov. 2002.

[24] I. J. Good, "Maximum entropy for hypothesis formulation, especially for multidimentional contingency tables," *Ann. Math. Statist.*, vol. 34, pp. 911–934, 1963.

[25] U. S. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley, CA: Univ. California Press, 1958.

[26] S. Haykin, *Nonlinear Methods of Spectral Analysis*. New York: Springer-Verlag, 1979.

[27] K. Hoffman, *Banach Spaces of Analytic Functions*. Englewood Cliffs, NJ: Prentice-Hall, 1962.

[28] R. E. Kalman, "Realization of covariance sequences," in *Proc. Toeplitz Memorial Conf. (1981)*, Tel-Aviv, Israel, 1981.

[29] A. Kavčić and J. M. F. Moura, "Matrices with banded inverses: Inversion algorithms and factorization of gauss-markov processes," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1495–1509, July 2000.

[30] H. Kimura, "Positive partial realization of covariance sequences," in *Modeling, Identification and Robust Control*, C. I. Byrnes and A. Lindquist, Eds. Amsterdam, The Netherlands: North-Holland, 1987, pp. 499–513.

[31] S. Kullback, *Information Theory and Statistics*, 2nd ed. New York: Dover, 1968.

[32] A. Lindquist and G. Picci, "Canonical correlation analysis, approximate covariance extension, and identification of stationary time series," *Automatica*, vol. 32, pp. 709–733, 1996.

[33] R. Nagamune, "A robust solver using a continuation method for nevanlinna-pick interpolation with degree constraint," *IEEE Trans. Automat. Contr.*, vol. 48, pp. 113–117, Jan. 2003.

[34] A. N. Amini and T. T. Georgiou, "Statistical analysis of state covariance subspace estimation methods," in *Proc. 41st IEEE Conf. Decision and Control*, Las Vegas, NV, Dec. 2002, pp. 2633–2638.

[35] M. A. Nielsen and J. L. Chuang, *Quantum Computation & Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[36] S. U. Pillai, T. I. Shim, and D. C. Youla, "A new technique for ARMA-system identification and rational approximation," *IEEE Trans. Signal Processing*, vol. 41, pp. 1281–1304, Mar. 1993.

[37] A. Renyi, *Probability Theory*. Amsterdam, The Netherlands: North-Holland, 1970.

[38] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1974.

[39] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[40] J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*. Providence, RI: Amer. Math. Soc., 1956, vol. 20, American Mathematical Society Colloquium Publications.