Likelihood Analysis of Power Spectra and Generalized Moment Problems

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

***Abstract*—We develop an approach to the spectral estimation that has been advocated by [A. Ferrante *et al.*, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2561–2575, Oct. 2012.] and, in the context of the scalar-valued covariance extension problem, by [P. Enqvist and J. Karlsson, "Minimal itakura-saito distance and covariance interpolation," in *Proc. 47th IEEE Conf. Decision Control*, 2008, pp. 137–142]. The aim is to determine the power spectrum that is consistent with given moments and minimizes the relative entropy between the probability law of the underlying Gaussian stochastic process to that of a prior. The approach is analogous to the framework of earlier work by Byrnes, Georgiou, and Lindquist and can also be viewed as a generalization of the classical work by Burg and Jaynes on the maximum entropy method. In this paper, we present a new fast algorithm in the general case (i.e., for general Gaussian priors) and show that for priors with a specific structure the solution can be given in closed form.**

***Index Terms*—Maximum likelihood estimation, spectral analysis, method of moments.**

I. Introduction

CONSIDER a stationary, vector-valued, discrete-time, zero-mean, Gaussian stochastic process $\{y(t) \mid t \in \mathbb{Z}\}$, where $y(t) \in \mathbb{R}^m$, and $\mathbb{Z}$, $\mathbb{R}$ are the sets of integers and reals, respectively. We denote the corresponding probability law (on sample paths of the process) by $\mathcal{P}$ [3, Ch. 1] and the power spectral density, which we assume exists, by $\Phi(e^{i\theta}), \theta \in [0, 2\pi)$. Further, we assume that the stochastic process is nondeterministic in that the entropy rate is finite

$$\int_{-\pi}^{\pi} \log\det \Phi(e^{i\theta}) d\theta < \infty. \tag{1}$$

Manuscript received October 30, 2016; accepted February 9, 2017. Date of publication February 22, 2017; date of current version August 28, 2017. This work was supported in part by the NSF under Grant ECCS-1509387, in part by the AFOSR under Grant FA9550-12-1-0319, Grant FA9550-15-1-0045, and Grant FA9550-17-1-0435, in part by the Vincentine Hermes-Luh Chair, in part by the Swedish Research Council (VR), and in part by the Foundation for Strategic Research. Recommended by Associate Editor S. Yuksel.

T. T. Georgiou is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697 USA (e-mail: tryphon@uci.edu).

A. Lindquist is with the Department of Automation and the School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Department of Mathematics, KTH Royal Institute of Technology, Stockholm 114 28, Sweden (e-mail: alq@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2017.2672862

This is a fairly general class that includes, e.g., all stochastic processes with nonsingular rational power spectral densities. We study the basic problem to estimate Φ from sample-statistics of $\{y(t)\}$. Following [1], we view this problem in a large-deviations framework where a prior law $\mathcal{Q}$ is available, and where this law corresponds to a power spectral density Ψ with finite entropy rate. We postulate that available sample-statistics of the process are not consistent with the prior law $\mathcal{Q}$, and therefore, we seek the law $\mathcal{P}$ that is consistent with these statistics and is the closest such law to the prior in the sense of large deviations, which amounts to $\mathcal{P}$ being such that the Kullback–Leibler (KL) divergence [4] between $\mathcal{P}$ and $\mathcal{Q}$ is minimal. Our approximation problem was considered in [2] for the special case of comparing the Itakura–Saito distance between scalar-valued time series since, in fact, the Itakura–Saito distance between the corresponding power spectra is precisely the KL-divergence between the two laws [3, Ch. 10], [5]–[7].

The theme of the approach, namely, to obtain power spectra that are consistent with empirical statistics and optimal with respect to a suitable criterion, is a standard recurring theme in works going back to Burg [8]. The problem to obtain empirical statistics from data is discussed in [9]–[11] and will not be addressed in this paper. Statistics represent (generalized) moment constraints and, in the past 30 years, a rich theory emerged that made contact with analytic function theory and the classical moment problem, see [12]–[38] and the references therein. A detailed and rigorous exposition of related topics and ideas in signal processing is given in [39].

Initially, following Burg, early researcher works focused on the entropy rate (1) as such a suitable functional to analyze geophysical time series. This eventually became dominant in speech processing under the acronym LPC (Linear Predictive Coding) [40]. The entropy rate relates to the variance of one-step-ahead linear prediction and the problem reduces to solving a linear set of equations, the normal equations [41]. In the context of autoregressive modeling these are solved by the Levinson algorithm. It soon became apparent that Burg's method was a special case of the Itakura–Saito autocorrelation approach, which in turn amounted to minimization of the discrimination information between a nominal model and a prior in the sense of the KL-divergence between their probability laws.

Subsequent developments viewed spectral estimation as an inverse problem to achieve consistency with estimated statistics. Initial motivation was provided by a question of R.E.

Kalman to identify spectra of low complexity [42]. Early results were obtained using topological and homotopy methods and the complete parametrization of solutions with generic minimal degree was formulated in steps in [13], [43], and [44]. Subsequently, it was discovered that optimizers of weighted entropy-like functionals (KL-divergence between power spectra as well as various types of distance to priors [6], [7]) had a particularly nice structure; they were rational and had small dimension [16]–[34], [45]–[47]. In fact, it turned out that suitably specified weighted entropy functionals contained the precise degrees of freedom that were needed to efficiently parametrize and construct these generic minimal degree solutions [16], [18]–[20], [28]. The mathematical underpinnings of this latter theory were largely based on optimization and duality, and closed the circle to once again connect with the KL-divergence [22]. This work is similar in spirit and technique but differs substantially in the choice of criterion and interpretation.

More specifically, following [1], [2], we consider the KL-divergence between *Gaussian probability laws* of stochastic processes or, equivalently, the Itakura–Saito distance between their power spectra. The interpretation as well as the structure of optimizers have subtle differences from earlier constructions. For one thing, the use of the KL-divergence in this way has a very natural and appealing interpretation: The sought power spectra represent *the most likely statistical signature on the path space of a time series that is in agreement with the estimated sample statistics* (see Section II-A). The structure of solutions retains many of the attractive features of earlier works. In particular, it ensures reasonably good bounds on the dimensionality of modeling filters (see Remark 5).

A comparison of the contributions of our paper to those in [1] is in order. The authors of [1] consider a pair of dual optimization problems, which we present in Section VI. Numerical complications occur in the approach of [1] due to a redundancy in the dual problem. One of the main contributions of this paper is to remove this redundancy by expressing the dual functional in a natural coordinate system that is based on our results in Section III. The corresponding reformulation of the dual problem leads to the fast algorithm presented in Section IV. A second main contribution is to provide a solution in closed form for certain choices of prior. These are analogous to the autoregressive models that arise in the case of trigonometric moment problems and all-pole priors. The results are presented for multivariable time series and moment problems for the corresponding matricial power spectra.

Below, in Section II, we begin by discussing in some detail the motivation for choosing the particular functional to guide identifying suitable power spectra that reproduce sample statistics. We then explain how sample statistics impose moment constraints on sought power spectra. In Section III, we present a geometric framework for input-to-state filters that provides basic tools for building a fast algorithm to solve the basic estimation problem. Section IV gives the problem formulation and presents the main results. Section V provides a simple example and connections with the earlier literature. Proofs of the main results are given in Sections VI–VIII. In particular, Section VII is devoted to deriving the fast algorithm and Section VIII to deriving the closed-form solution, respectively. In the concluding Section IX we provide some final thoughts.

II. Preliminaries

A. Likelihood Framework

The rationale for the framework adopted herein has been used to justify maximum likelihood methods [48]–[50] and complements the original reasoning by E.T. Jaynes [51]–[53]. It can be presented as follows. If sample paths of a time series are drawn out of the given prior $\mathcal{Q}$, they have a small probability of giving rise to sample statistics that are not consistent with $\mathcal{Q}$. If that were to happen, and thereby, the sample paths represent a rare event, i.e., a departure from what is expected, one is motivated to seek out of the many possible sample-path distributions that are consistent with the observed statistics the one that is most likely. It is known that, asymptotically, the probability of rare events that suggest an (empirical) distribution $\mathcal{P}$ depends exponentially on the KL-divergence between the prior $\mathcal{Q}$ and $\mathcal{P}$ [54], [55]. There exists a vast literature on applications of such a large deviations viewpoint to system identification, primarily with emphasis on parameter estimations and stochastic approximation, see e.g., [56] and references therein.

The KL-divergence between two laws $\mathcal{P}$ and $\mathcal{Q}$ is

$$D(\mathcal{P}\|\mathcal{Q}) = \lim_{N\to\infty} \frac{1}{2N+1} D(\mathcal{P}|_{[-N,N]}\|\mathcal{Q}|_{[-N,N]}) \quad (2)$$

where $\mathcal{P}|_{[-N,N]}$ denotes the restriction of $\mathcal{P}$ to the subset of random variables

$$\{y(-N),\dots,y(-1),y(0),y(1),\dots,y(N)\}$$

and similarly for $\mathcal{Q}|_{[-N,N]}$. In turn, the KL-divergence between the finite-dimensional probability densities $p(y(-N),\dots,y(N))$ and $q(y(-N),\dots,y(N))$, corresponding to $\mathcal{P}|_{[-N,N]}$ and $\mathcal{Q}|_{[-N,N]}$, is

$$\int_{\mathbb{R}^{2N+1}} p\log(q/p)\; dy(-N)\cdots dy(N).$$

Provided both laws represent purely nondeterministic processes, as is assumed herein, the limit in (2) exists. Using Szegö-Wiener-Masani's formula (see e.g., [57, Lemma 5.1], [58, formula (E.12)], [59, Th. 11.3.5]), $D(\mathcal{P}\|\mathcal{Q})$ can be expressed in terms of the corresponding power spectral densities as follows:

$$\begin{aligned} D(\mathcal{P}\|\mathcal{Q}) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left(\Phi\Psi^{-1} - \log\Phi\Psi^{-1} - I\right) d\theta \\ &=: \mathbb{D}(\Phi\|\Psi) \end{aligned} \quad (3)$$

where $\operatorname{tr}(\cdot)$ denotes trace. Since $\mathcal{P}$ is completely specified by Φ we only need to determine Φ, based of course on empirical statistics. Thus, we are interested in determining a power spectral density Φ that is consistent with given statistics and minimizes $\mathbb{D}(\Phi\|\Psi)$ for a given power spectrum Ψ. The precise formulation of the problem requires expressing statistics in terms of power spectra, which is done next. The problem is stated precisely in Section IV.

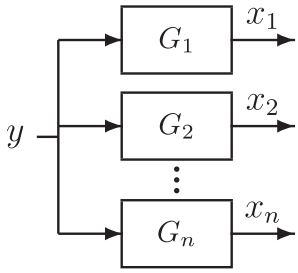

Fig. 1. Bank of filters.

B. Filter Banks and Statistics

Time-series represent samples of a stochastic process, and available statistics consist of sample covariances. We now explain the setting and nature of the covariance data.

In time-series analysis as well as in antenna array processing it is customary to assume that recorded data are scaled by a frequency-dependent vector/matrix-valued gain $G(e^{i\theta})$ where the frequency θ corresponds to time, space, angle, or even a vector-valued combination, see e.g., [39, Ch. 6], [60], [61], [27, Sec. II-B]. For instance, a window of observations $\{y(k),\, y(k-1), \ldots,\, y(k-n)\}$ of a time-series can be thought as the vectorial output of a "tapped delay line" represented by the vector-valued Fourier vector $\begin{bmatrix} 1 & e^{-i\theta} & \ldots & e^{-in\theta} \end{bmatrix}'$, i.e., the Fourier vector is the transfer function of the tapped delay line. Likewise, in the array processing literature, a model of an equispaced array of $n+1$ omnidirectional sensors registering signals that are emitted from afar is again the same Fourier vector [39, Sec. VI]. Such a vector-valued gain G, for general arrays, is often referred to as the *array manifold* and can be thought as a bank of filters that capture the relative dependence of the sensor outputs to signals from afar (see Fig. 1). Often, for a large equispaced array of sensors, a smaller output is selected that corresponds to G being a linear combination of Fourier components (beamspace techniques)

$$G(e^{i\theta}) = M \begin{bmatrix} 1 \\ e^{-i\theta} \\ \vdots \\ e^{-i(n-1)\theta} \end{bmatrix} \tag{4}$$

for a suitable matrix M. Other times, processing of time series or sensor-array data involves a suitably designed bank of filters $G_k(e^{i\theta})$, $k = 1, 2, \ldots, n$,

in which case

$$G(e^{i\theta}) = \begin{bmatrix} G_1(e^{i\theta})\; G_2(e^{i\theta}),\; \ldots,\; G_n(e^{i\theta}) \end{bmatrix}'$$

with $\{y(t)\}$ the common input and general dynamics, see e.g. [21], [62]. The filters may also encapsulate attenuation from the coordinate θ of "sources" generating $\{y(t)\}$ to the respective outputs of sensor array (cf. [27, Sec. II]). In all these cases, it is natural to estimate covariance of the vectorial time series

$$x(t) = \begin{bmatrix} x_1(t)\; x_2(t),\; \ldots,\; x_n(t) \end{bmatrix}'.$$

This is typically the form of available statistics that we consider henceforth.

We assume that G is a square-integrable, stable $n \times m$ transfer function. Then, the n-dimensional output process $\{x(t) \mid t \in \mathbb{Z}\}$ assumes a representation as a stochastic integral

$$x(t) = \int_{-\pi}^{\pi} e^{-it\theta} G(e^{i\theta}) \underbrace{W(e^{i\theta})\, d\hat{w}(\theta)}_{d\hat{y}(\theta)}$$

where $\hat{w}$ is a Wiener process such that $\mathrm{E}\{d\hat{w} d\hat{w}^*\} = I d\theta/2\pi$. Here, I is the identity matrix, $\mathrm{E}\{\ \}$ is the expectation operator, and W is a (minimum-phase) spectral factor of Φ, i.e., $W(e^{i\theta})W(e^{i\theta})^* = \Phi(e^{i\theta})$, and therefore, $d\hat{y}$ is the stochastic Fourier transform of y; see e.g., [59, Ch. 3]. It follows that the $n \times n$ covariance of the (zero-mean) vectorial output $x(t)$ is

$$\Sigma := \mathrm{E}\{x(t)x(t)'\} = \int G \Phi G^* \tag{5}$$

where, for economy of notation, we have suppressed the limits of integration and the normalized Lebesgue measure $d\theta/2\pi$, i.e., $\int$ denotes $\int_{-\pi}^{\pi} \frac{d\theta}{2\pi}$. The value Σ represents a matricial moment constraint on Φ. The problem that we consider below is, given G and Σ, to determine suitable Φ satisfying (5).

C. Input-to-State Filters

A special case of a filter bank of great interest is when this represents an input-to-state (stable) linear system

$$x(t) = Ax(t-1) + By(t),\ t \in \mathbb{Z} \tag{6}$$

where $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$. In that case, the transfer function of the filter bank is

$$G(z) = z(zI - A)^{-1}B. \tag{7}$$

Throughout, we assume that all the eigenvalues of A are located in the open unit disc. Then

$$\begin{aligned} G(z) &= (I - z^{-1}A)^{-1}B \\ &= B + ABz^{-1} + A^2Bz^{-2} + A^3Bz^{-3} + \ldots \end{aligned}$$

for all z such that $|z| > 1$. Throughout, to insure that the complete state space is being reached and to avoid trivialities we assume that (A, B) is a reachable pair, i.e.

$$\text{rank}\,[B, AB, \cdots, A^{n-1}B] = n$$

and that B is full column rank. The use of such filter banks is the basis of a tunable method of spectral analysis that was introduced in [21] and is referred to as THREE.

The input-to-state structure in (7) encompasses Fourier vectors where $G_k(z) := z^{-(k-1)}$, $k = 1, 2, \ldots, n$, in that case

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{8}$$

and the $n \times n$ state covariance is Toeplitz, i.e.

$$\Sigma := E\{x(t)x(t)'\} = \begin{bmatrix} c_{k-\ell} \end{bmatrix}_{k,\ell=1}^{n} \tag{9}$$

where $c_k := \mathrm{E}\{y(t+k)y(t)\}$. Identifying a power spectral density Φ that is consistent with Σ and the process model is precisely the problem that underlies subspace identification [59] and coincides with the classical "covariance extension" or "trigonometric moment" problem.

On the other hand, first-order filters $G_k(z) := \frac{z}{z-p_k}$, $k = 1, 2, \dots, n$, (with $p_k \neq p_\ell$ for $k \neq \ell$) lead to

$$A = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_n \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{10}$$

and a state covariance matrix Σ that has the structure of a Pick matrix; see [22].

Finally, it is also seen that (7) is of the form (4) where $M = [B, AB, \dots]$. This matrix is finite when A is nilpotent corresponding to "moving average" dynamics.

III. Geometry of Input-to-State Filters

The (rational) input-to-state structure of $G(z)$ in (7) imposes structural algebraic constraints on the covariance of $x(t)$. In addition to positive definiteness, Σ is completely characterized by belonging to the range of the integral operator

$$\Gamma \,:\, \Phi \mapsto \Sigma = \int G\Phi G^*. \tag{11}$$

This is a linear operator that takes $m \times m$ integrable matrix-valued functions Φ on the unit circle to symmetric matrices Σ.

The range of Γ admits an algebraic characterization. In fact, it is shown in [63] that a symmetric $n \times n$ matrix Σ belongs to $\mathrm{range}(\Gamma)$ if and only if

$$\Sigma - A\Sigma A' = BH + H'B' \tag{12}$$

for some $m \times n$ matrix H. Equivalently,

$$\mathrm{rank} \begin{bmatrix} \Sigma - A\Sigma A' & B \\ B' & 0 \end{bmatrix} = \mathrm{rank} \begin{bmatrix} 0 & B \\ B' & 0 \end{bmatrix} \tag{13}$$

where 0 denotes a zero-matrix of appropriate size, is necessary and sufficient for solvability of (12). Moreover, there is a coercive, continuous spectral density Φ satisfying the generalized moment condition (5) if and only if Σ is positive definite[1] and satisfies (12) or, the equivalent condition (13).

The adjoint operator Γ^* maps symmetric matrices into $m \times m$ integrable Hermitian matrix-valued functions on the unit circle, namely

$$\Gamma^* \,:\, \Lambda \mapsto G^*\Lambda G.$$

The inner product in these two spaces, symmetric matrices and integrable Hermitian matrix-valued functions on the unit circle, relate as

$$\begin{aligned} \langle \Lambda, \Sigma \rangle &:= \mathrm{tr}(\Lambda\Sigma) \\ &= \mathrm{tr} \int G^*\Lambda G\Phi \\ &=: \langle G^*\Lambda G, \Phi \rangle. \end{aligned}$$

We also consider the operator

$$\Theta \,:\, H \mapsto \Delta = BH + H'B'$$

which maps $\mathbb{R}^{m\times n}$ to symmetric $n \times n$ matrices and its adjoint

$$\Theta^* \,:\, \Delta \mapsto 2B'\Delta. \tag{14}$$

We are interested in nonredundant representations of $\mathrm{range}(\Gamma)$ and $\mathrm{range}(\Gamma^*)$ by identifying the minimal degrees of freedom in suitable matrix representations. The first proposition deals with $\mathrm{range}(\Gamma)$.

Proposition 1: The map

$$\Sigma \mapsto H = (B'B)^{-1}\left[B'(\Sigma - A\Sigma A') - YB'\right] \tag{15a}$$

where Y is the symmetric solution of the Lyapunov equation

$$(B'B)Y + Y(B'B) = B'(\Sigma - A\Sigma A')B \tag{15b}$$

establishes a bijective correspondence between $\Sigma \in \mathrm{range}(\Gamma)$ and $H \in \mathrm{range}(\Theta^*)$.

Proof: Set $\Delta := \Sigma - A\Sigma A'$. Since we have $\Sigma \in \mathrm{range}(\Gamma)$

$$\Delta = BH + H'B' \tag{16}$$

can be solved for $H \in \mathbb{R}^{m\times n}$ and $\Delta = \Theta(H)$. We seek a particular solution of minimal Frobenius norm

$$\|H\|_{\mathrm{F}} := \sqrt{\mathrm{tr}(HH')}.$$

Then, this solution will be in $\mathrm{range}(\Theta^*)$, a fact that will be verified below. The Lagrangian of the problem is

$$\mathrm{tr}(HH') + 2\mathrm{tr}(\Lambda BH) - \mathrm{tr}(\Lambda\Delta)$$

where $\Lambda = \Lambda'$ is the symmetric matrix-valued Lagrange multiplier. It follows that the unique optimal solution is of the form

$$H = B'\Lambda \tag{17}$$

and therefore, $H \in \mathrm{range}(\Theta^*)$ in view (14). Then, $HB = B'\Lambda B =: Y$ is symmetric. Further, it satisfies the Lyapunov equation

$$(B'B)Y + Y(B'B) = B'\Delta B \tag{18}$$

as can be seen by premultiplying (16) by B' and postmultiplying by B. Since B has full column rank by assumption, the eigenvalues of $B'B$ are positive and (18) has a unique solution Y. By premultiplying (16) by B' we can now solve for

$$H = (B'B)^{-1}\left(B'\Delta - YB'\right). \tag{19}$$

Finally, suppose that (12) has two solutions H_1 and H_2 in $\mathrm{range}(\Theta^*)$. Then

$$H_1 - H_2 \in \ker\Theta = (\mathrm{range}(\Theta^*))^\perp$$

and hence $H_1 = H_2$, proving uniqueness.

[1]The case where Σ is only nonnegative definite is discussed fully in [64]. In that case the spectral content may correspond to a singular spectral measure.

The essence is that (12) has many solutions in general when $m > 1$. In that case, Θ has a nontrivial null space, and Proposition 1 provides the solution to (12) of minimal Frobenius norm.

The next proposition deals with $\mathrm{range}(\Gamma^*)$. Since the orthogonal complement of the range of Γ is the null space of Γ^*, elements in $\mathrm{range}(\Gamma^*)$ can always be written in the form $G^*\Lambda G$ where $\Lambda \in \mathrm{range}(\Gamma)$.

Proposition 2: The map

$$G^*\Lambda G \mapsto X = MB \tag{20a}$$

where M is the unique solution of the Lyapunov equation

$$M = A'MA + \Lambda \tag{20b}$$

establishes a bijective correspondence between $G^*\Lambda G \in \mathrm{range}(\Gamma^*)$ and $X \in \mathrm{range}(\Theta^*)$.

Proof: We first note that the dimension of $\mathrm{range}(\Gamma^*)$, which coincides with the dimension of $\mathrm{range}(\Gamma)$, is equal to the dimension of $\mathrm{range}(\Theta^*)$ by Proposition 1. Since M is symmetric, it also follows that $X' = B'M \in \mathrm{range}(\Theta^*)$. Thus, in order to establish that the correspondence $G^*\Lambda G \mapsto X'$ is a bijection, it suffices to prove that $X' = 0$ only when $\Lambda = 0$. To see this note that, since $AG(z) = z\big(G(z) - B\big)$, (12) yields

$$\begin{aligned} G^*\Lambda G &= G^*MG - G^*A'MAG \\ &= G^*MG - [G-B]^*M[G-B] \\ &= G_0^*X + X'G_0 \end{aligned} \tag{21}$$

with G_0 given

$$\begin{aligned} G_0(z) &:= G(z) - \tfrac{1}{2}B = \tfrac{1}{2}B + A(zI-A)^{-1}B \\ &= \tfrac{1}{2}B + ABz^{-1} + A^2Bz^{-2} + A^3Bz^{-3} + \dots. \end{aligned} \tag{22}$$

But, since $\Lambda \in \mathrm{range}(\Gamma)$, $G^*\Lambda G = 0$ only when $\Lambda = 0$. Thus, $X = 0$ implies that $\Lambda = 0$ and this completes the proof. ■

IV. Main Results

We are now in a position to formulate the main problem that we consider. As noted earlier this problem was first formulated and studied in [1].

Problem 1: Given an $m \times m$ matrix-valued power spectral density Ψ, and given the parameters A, B of the input-to-state filter (filter bank) in (6) and the covariance Σ of the state process $x(t)$, determine

$$\hat{\Phi} \in \mathrm{argmin}\{\mathbb{D}(\Phi\|\Psi) \mid \text{ such that (5) holds}\}.$$

We provide a solution to this problem under fairly general conditions on the prior spectral density Ψ, namely,

i) Ψ is coersive on the unit circle, and
ii) $\Psi(e^{i\theta})^{-1}$ is Lipschitz continuous in $\theta \in [-\pi, \pi]$.

In the theorem below we describe the structure of solutions. The expressions we give provide an alternative to those in [1] and require fewer variables in general. This nonredundant structure of solutions is analogous to the reduction in the number of variables enabling the fast algorithms for Kalman filtering in [65].

As in our previous work on the moment problem, e.g., [18], [20], [21], solving Problem 1 reduces to convex optimization. With G_0 given by (22), the optimization criterion is the strictly convex functional

$$\begin{aligned} \mathbf{J}(X) = \mathrm{tr}\Big\{ &(HX + X'H') \\ &- \int \log\big(\Psi^{-1} + G_0^*X + X'G_0\big)\Big\} \end{aligned} \tag{23}$$

defined on the open set $\mathcal{X}_+$ of matrices $X \in \mathbb{R}^{n\times m}$ such that $X' \in \mathrm{range}(\Theta^*)$, i.e., $B'X$ is symmetric, and

$$Q(z) := \Psi(z)^{-1} + G_0(z)^*X + X'G_0(z) \tag{24}$$

is positive definite at each point $z = e^{i\theta}$ on the unit circle.

Theorem 3: Let Σ be a symmetric, positive definite $n \times n$ matrix in the range of Γ, and let H be given by (15). Suppose that the prior spectral density Ψ satisfies conditions (i and ii) above. Then Problem 1 has the unique solution

$$\hat{\Phi} = \hat{Q}^{-1} \tag{25a}$$

where

$$\hat{Q} = \Psi^{-1} + G_0^*\hat{X} + \hat{X}'G_0 \tag{25b}$$

for some $\hat{X} \in \mathcal{X}_+$.

The matrix $\hat{X}$ is the unique minimizer of the functional $\mathbf{J}(X)$, and it is also the unique solution of the stationarity condition

$$\int \big(\Psi^{-1} + G_0^*X + X'G_0\big)^{-1}G_0^* = H. \tag{26}$$

Remark 1: In particular, conditions (i and ii) on the prior are satisfied by power spectra Ψ that are rational and have nonsingular determinants on the unit circle. In the case where Ψ is rational, the solution to Problem 1 is also rational, and thereby corresponds in general to an autoregressive moving-average (ARMA) model. On the other hand, Ψ can equally well be taken to be nonrational as this often arises in physical problems, e.g., when representing various types of scattering interference. ■

The solution provided in the theorem can be obtained numerically by a Newton method to compute the minimizer of $\mathbf{J}$. To this end, we compute the gradient

$$\frac{1}{2}\frac{\partial \mathbf{J}}{\partial X} = H - \int Q^{-1}G_0^* \tag{27a}$$

where Q is given by (24), and the Hessian

$$\frac{1}{2}\mathcal{H}(X) = \int G_0Q^{-2}G_0^* > 0. \tag{27b}$$

The positivity of the Hessian indeed shows that the functional $\mathbf{J}$ is strictly convex. A possible starting point is $X = 0$.

Remark 2: Unlike the situation in [1], we have explicit expressions (27) for the Hessian and the gradient in the parameter space $\mathcal{X}_+$. Therefore, we can apply Newton's method directly, whereas in [1, Sec. VI-A] the search direction needs to be determined implicitely to keep the iteration point in the

space $\mathrm{range}(\Gamma)$. Moreover, the variable X in the above theorem and expressions (27) belongs to $\mathcal{X}_+$, which has dimension $nm - \frac{1}{2}(m^2 - m)$, since $X' \in \mathrm{range}(\Theta^*)$. This should be compared with the $\frac{1}{2}n(n+1)$ variables in the dual functional (34) used in [1]. This is the reason for referring to our procedure as a fast algorithm. ■

In the next theorem, we consider the special case where the prior Ψ has the form $\Psi = (G^*\Lambda_0 G)^{-1}$. Then the solution to Problem 1 can be given in closed form.

Theorem 4: Let Σ be a positive definite $n \times n$ matrix in the range of Γ, and suppose that the prior Ψ is given by

$$\Psi = (G^*\Lambda_0 G)^{-1}. \tag{28}$$

Then Problem 1 has the unique solution

$$\hat{\Phi} = \left(G^*\hat{\Lambda}G\right)^{-1} \tag{29}$$

where

$$\hat{\Lambda} := \Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}$$

and does not depend on Λ_0.

The proofs of Theorems 3 and 4 are given in Sections VII and VIII, respectively.

Remark 3: It is interesting to point out that the solution $\hat{\Phi}$ to Problem 1 shares the same zeros as the prior Ψ. To see this note that at any value on the complex plane where Ψ becomes singular, $\hat{Q}$ becomes infinity along suitable direction, and therefore, $\hat{\Phi}$ becomes singular as well. This property is also present in solutions to moment problems that minimize alternative entropy functionals and has been explored in our earlier work. It is quite instructive to consider Problem 1 in the scalar case ($m = 1$). Then the optimal solution takes the form

$$\hat{\Phi}(e^{i\theta}) = \frac{\Psi(e^{i\theta})}{1 + 2\Psi(e^{i\theta})\mathrm{Re}\{G_0(e^{i\theta})^*\hat{X}\}}. \tag{30}$$

Any zeros of the prior Ψ will, therefore, be zeros also of $\hat{\Phi}$. However, in the special case of rational Ψ, the dimension of modeling filters corresponding to $\hat{\Phi}$ is enlarged as compared with alternative formulations in our earlier works, e.g., [20], [21]. ■

Remark 4: Since the closed-form solution (29) does not depend on Λ_0, we may in particular choose $\Lambda_0 = I$. Then, in the important case when $G_k(z) := z^{-(k-1)}$, $k = 1, 2, \ldots, n$, we have $\Psi = I$, leading to an autoregressive (maximum-entropy) model. ■

Remark 5: Going back to [42], the original motivation was to identify and characterize solutions to moment problems having low degree. It is instructive to consider the scalar trigonometric moment problem with data (8) and (9), that is, the problem to match the n covariance samples $\{c_0, c_1, \ldots, c_{n-1}\}$ with a rational power spectrum Φ, in the sense that

$$c_k = \int e^{ik\theta}\Phi \text{ for } k = 0, 1, \ldots, n-1$$

holds, or, equivalently, (5) holds for the $n \times n$ covariance matrix Σ. There is a generic set of covariance samples (i.e., a set with an open interior) for which the minimal degree solution has spectral factors of degree $n - 1$ [43]. (For a more general result of this type; see [66, Th. 2.2].) The family of all power spectra with the same dimensionality can be parametrized by a set of arbitrarily selected $n - 1$ spectral zeros (i.e., zeros of the corresponding minimum-phase spectral factor)—existence of power spectra corresponding to each such choice was shown in [13], [43] and uniqueness was shown in [44]. Likewise, in the case of m-vector valued time series where an $n \times n$ covariance matrix Σ is available, the family of generically minimal degree solutions has spectral factors of degree $n - m$, parametrized accordingly for a choice of spectral zeros [67, Sec. IV and Corollary 2]). On the other hand, a direct approach of constructing solutions based on the THREE framework gives a family of solutions with spectral factors of degree n [27, Sec. IV-B] (instead of the generic minimum $n - m$ in [67]) likewise parametrized by a suitable choice of spectral zeros. The current framework allows constructing solutions (29) with spectral factors of degree n only when the zero-structure is trivial (i.e., identical to the eigenvalues of the matrix A), while in general the best bound one can provide from (25) for the dimension of spectral factors is $n + \frac{1}{2}\times$(degree of Ψ); cf. [1, Sec. IV]. ■

V. Simple Example

In this example, we consider as data a particular matrix Σ that originates as the state covariance of a filter (6). Evidently, there is a plethora of power spectra Φ that are all solutions to the moment equation (5), i.e.

$$\Sigma = \int G\Phi G^*.$$

In the absence of particular knowledge about the power spectrum of the underlying process one would normally assume the uniform prior leading to the "maximum entropy" solution. Instead, if one begins with a reasonably good prior, the power spectrum that is closest and agrees with the particular Σ seems a more logical alternative. The theory in this paper and in the papers [1], [2] deal precisely with this situation. Thus, in this section we present and compare the two power spectra, one labeled $\hat{\Phi}$ that is based on a prior Ψ with low-pass character and the maximum entropy solution Φ_{ME}. The covariance Σ is intentionally chosen to correspond to a power spectrum that has a triple zero at $\theta = \pi$. As a result, although both power spectra $\hat{\Phi}$ and Φ_{ME} are consistent with the covariance data, the former is closer to the low pass character of the generating spectrum by virtue of a similar character of the prior. The point of the example is not compare the "performance" of the two methods since either matches the moments, but rather their dramatically different behavior and to suggest how this can be influenced by the availability of a prior.

We consider the case where Σ is a Toeplitz matrix as given in (9) with covariance lags $c_k := \mathrm{E}\{y(t+k)y(t)\}$ of a scalar stationary process y. Then G is given by (4) with $M = I$. Moreover, A and B are given by (8), and hence $B'(\Sigma - A\Sigma A) = (c_0, c_1, \ldots, c_{n-1})$ and $B'B = 1$. Consequently, it follows from (18) that $Y = \frac{1}{2}c_0$ and from (19) that $H = (\frac{1}{2}c_0, c_1, \ldots, c_{n-1})$.

Then, setting $X' := (q_0, q_1, \ldots, q_{n-1})$, we have

$$HX + X'H' = \langle c, q \rangle := \sum_{k=-(n-1)}^{n-1} c_k q_k$$

and

$$Q(e^{i\theta}) = \Psi(e^{i\theta})^{-1} + \sum_{k=-(n-1)}^{n-1} q_k e^{ik\theta}. \tag{31a}$$

Problem 1 then amounts to minimizing

$$\mathbf{J}(q) = \langle c, q \rangle - \int \log Q \tag{31b}$$

over all $q := (q_0, q_1, \ldots, q_{n-1})$ such that $Q(e^{i\theta}) > 0$ for all θ. In this notation the stationarity condition (26) becomes

$$\int_{-\pi}^{\pi} e^{ik\theta} Q^{-1} \frac{d\theta}{2\pi} = c_k, \quad k = 0, 1, \ldots, n-1. \tag{31c}$$

Remark 6: It is interesting to compare the functional $\mathbf{J}(q)$ and the form of solution above to those in the framework of, e.g., [18], [20], [21]. There, the corresponding functional is $\mathbf{J}(q) = \langle c, q \rangle - \int \Psi \log Q$ instead of (31b), with

$$Q(e^{i\theta}) = \sum_{k=-(n-1)}^{n-1} q_k e^{ik\theta}$$

and moment conditions $\int e^{ik\theta} \frac{\Psi}{Q} \frac{d\theta}{2\pi} = c_k$, $k = 0, 1, \ldots, n-1$, instead of (31). We see that the present framework is analogous to the maximum-entropy solution in these earlier works except for the absence of $\Psi(e^{i\theta})^{-1}$ in the corresponding expression for Q, which is traded off with the direct presence of Ψ in functional and the stationarity conditions. The optimal solution is $\hat{\Phi}(e^{i\theta}) = \Psi(e^{i\theta}) / \sum q_k e^{ik\theta}$ in this case instead of $\hat{\Phi}(e^{i\theta}) = \Psi(e^{i\theta}) / 1 + \Psi(e^{i\theta}) \sum q_k e^{ik\theta}$ in our present framework. ■

We proceed with our numerical example. To this end, we select

$$\Phi(z) = |(z+1)(z^{-1}+1)|^3 \tag{32}$$
$$= z^{-3} + 6z^{-2} + 15z^{-1} + 20 + 15z + 6z^2 + z^3$$

that corresponds to a moving average filter with transfer function $W(z) = 1 + 3z^{-1} + 3z^{-2} + z^{-3}$. In Fig. 2, we first compare the "true" (or, rather, reference) power spectral density Φ in (32), evaluated at $z = e^{i\theta}$ for $\theta \in [0, \pi]$, with a prior $\Psi = 10(1 + 0.9\cos(\theta)(1 + 0.9\cos(\theta))^2)$ that is selected to have a low pass characteristic. We seek to match eight moments, namely, $c = (20, 15, 6, 1, 0, 0, 0, 0)$. Next, in Fig. 3, we compare Φ with the optimal solution $\hat{\Phi}$ to Problem 1 for the given Ψ. Finally, in Fig. 4, we compare Φ with the solution corresponding to the choice $\Psi = 1$. The power spectral density obtained in this way, using either the Newton algorithm based on Theorem 3 or the closed-form expression in Theorem 4, is an *all-pole* power spectrum that agrees with the given moments, i.e., an autoregressive (AR) model. In contrast, $\hat{\Phi}$ corresponds to an ARMA model.

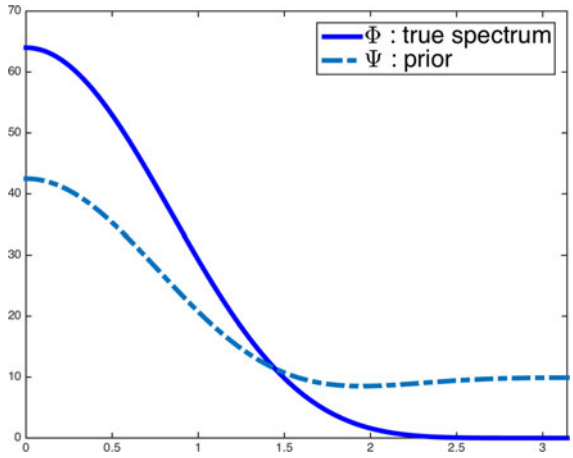

Fig. 2. Reference spectrum (solid line) versus prior (dashed).

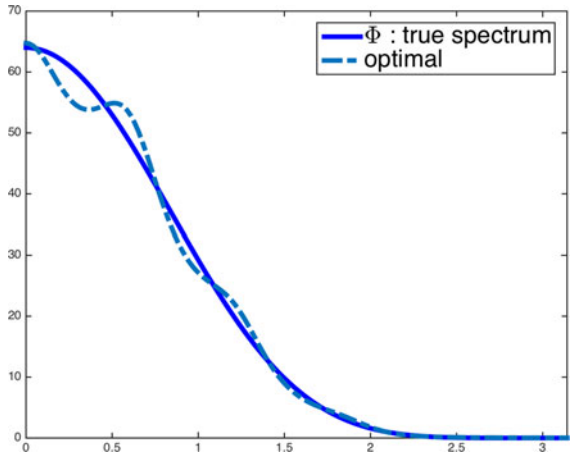

Fig. 3. Reference spectrum (solid line) versus optimal (dashed).

It is interesting to observe the oscillatory character of the all-pole power spectral density. In contrast, the use of a prior with a low pass character alleviates the oscillations (see Fig. 3).

Remark 7: We stress again that the "true" spectral density plays a subordinate role in this example. Indeed, if instead the maximum-entropy solution in Fig. 4 was the "true" spectral density, it would produce exactly the same covariance data as used in the example. In that case, the ARMA solution in Fig. 3, constructed with the now misleading prior, would entirely miss the ripples in the true spectrum. On the other hand, if we were told that the data originates from an all-pole spectrum, and thereby used an antireflection (AR) spectrum as a prior instead, the solution to the optimization problem would produce the unique AR power spetrum shown in Fig. 4, which is consistent with the given covariance data.

VI. Dual Problem and the Form of the Minimizer

Suppose that Σ belongs to the range of the operator Γ, defined by (11). Then, Problem 1 amounts to minimizing

$$\mathbb{D}(\Phi \| \Psi) = \frac{1}{2} \int \mathrm{tr}\left(\Phi\Psi^{-1} - \log \Phi + \log \Psi - I\right) \tag{33a}$$

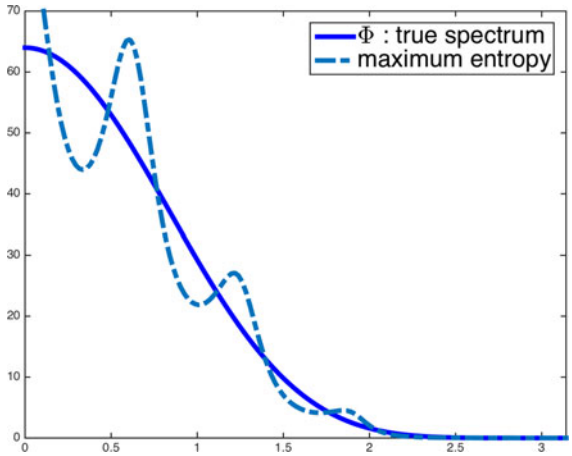

Fig. 4. Reference spectrum (solid line) versus ME spectrum (dashed).

over all spectral densities Φ satisfying the moment condition

$$\Sigma = \int G\Phi G^*. \tag{33b}$$

Proceeding along the lines of [22], it was shown in [1] that the dual of (33) is the problem to minimize

$$\mathbb{J}(\Lambda) = \mathrm{tr}\left\{\Lambda\Sigma - \int \log Q\right\} \tag{34a}$$

over all real, symmetric $n \times n$ matrices Λ in the range of Γ such that

$$Q(z) := \Psi(z)^{-1} + G(z)^*\Lambda G(z) \tag{34b}$$

is positive on the unit circle. For the convenience of the reader, we also review some steps in the proof in our present notation.

We denote the class of feasible Λ by $\mathcal{L}_+$, i.e.

$$\mathcal{L}_+ = \{\Lambda \in \mathrm{range}(\Gamma) \mid \Lambda' = \Lambda;\ Q(e^{i\theta}) > 0,\ \forall\theta\}.$$

We note in passing that the rationality of G is not needed at this point; in fact, an interesting example with $G(e^{i\theta}) = [1,\ e^{i\theta},\ e^{i\sqrt{2}\theta}]'$ is motivated in the context of sensor array processing in [27].

The Lagrangian for the problem above becomes

$$\begin{aligned} L(\Phi,\Lambda) &= \mathbb{D}(\Phi\|\Psi) + \mathrm{tr}\left\{\Lambda\left(\int G\Phi G^* - \Sigma\right)\right\} \\ &= -\mathrm{tr}(\Lambda\Sigma) + \int \mathrm{tr}\left\{\Phi(\Psi^{-1} + G^*\Lambda G)\right. \\ &\qquad \left. - \log\Phi + \log\Psi - I\right\} \end{aligned}$$

where Λ is a symmetric $n \times n$ matrix of Lagrange multipliers. Since $\mathrm{tr}\left\{\Lambda\left(\int G\Phi G^* - \Sigma\right)\right\}$ is simply the inner product of Λ with elements in the range of Γ, we can restrict Λ to the same space, and therefore, assume that

$$\Lambda \in \mathrm{range}(\Gamma).$$

The function $\Phi \mapsto L(\Phi,\Lambda)$ is strictly convex for each Λ such that Q, defined by (34b), is positive semidefinite on the unit circle. If Q fails to be positive semidefinite, $L(\Phi,\Lambda)$ can be made arbitrarily small for some Φ, and hence such a Λ is not a candidate in the dual problem. Hence, we may restrict Λ to the class $\mathcal{L}_+$. Setting the directional derivative

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

equal to zero, we obtain

$$\Phi = (\Psi^{-1} + G^*\Lambda G)^{-1} \tag{35}$$

which inserted into the Lagrangian yields the dual functional

$$\begin{aligned} \varphi(\Lambda) &= -\mathrm{tr}(\Lambda\Sigma) \\ &\quad + \int \mathrm{tr}\left\{\log(\Psi^{-1} + G^*\Lambda G) + \log\Psi\right\} \\ &= -\mathbb{J}(\Lambda) + \int \mathrm{tr}\log\Psi. \end{aligned}$$

Since this dual functional should be maximized, the dual problem is equivalent to minimizing $\mathbb{J}$ over all $\Lambda \in \mathcal{L}_+$. It was shown in [1] that this problem has a unique solution. This problem differs from the one in [22] in that the prior Ψ in [22] does not occur in Q but instead multiplies $\log Q$. Unlike the situation in [22], $\mathrm{tr}(\Lambda\Sigma)$ might be negative in the present setting that complicates the analysis somewhat. Nevertheless, the functional $\mathbb{J}$ is bounded from below, as stated next, a fact that will be used in Section VII.

Lemma 5: If Σ belongs to the range of Γ, then the functional $\mathbb{J}$ is bounded from below.

Proof: The condition that Σ belongs to the range of Γ ensures the existence of a spectral density Φ_0 satisfying (5). Then, in view of the construction above, $\varphi(\Lambda) \le L(\Phi_0,\Lambda) = \mathbb{D}(\Phi_0\|\Psi)$ or equivalently

$$\mathbb{J}(\Lambda) \ge \int \mathrm{tr}\log\Psi - \mathbb{D}(\Phi_0\|\Psi)$$

which establishes the required bound.

VII. Redux on the Dual Problem: The Fast Algorithm

Next we turn to the proof of Theorem 3. One of the difficulties dealing with the dual problem in Section VI is the redundancy introduced by the integral operator Γ, which has the consequence that only the part of Λ belonging to the range of Γ affects the value of $\mathbb{J}(\Lambda)$. To remove this redundancy we reformulate the problem by defining $\mathbb{R}^{n\times m}$ matrix-valued variable

$$X = MB \tag{36}$$

where M is the unique solution of the Lyapunov equation

$$M = A'MA + \Lambda \tag{37}$$

and $\Lambda \in \mathrm{range}(\Gamma)$ is the matrix-valued variable in the dual problem in Section VI. By Proposition 2, there is a one–one correspondence between Λ and X. In view of (21)

$$G(z)^*\Lambda G(z) = G_0(z)^*X + X'G_0(z)$$

where G_0 is given by (22). Therefore, (34b) takes the form

$$Q(z) = \Psi(z)^{-1} + G_0(z)^* X + X' G_0(z). \tag{38}$$

Moreover, in view of (37) and (12)

$$\begin{aligned}\mathrm{tr}(\Lambda\Sigma) &= \mathrm{tr}(M\Sigma) - \mathrm{tr}(MA\Sigma A') \\ &= \mathrm{tr}(MBH) + \mathrm{tr}(B'MH') \\ &= \mathrm{tr}(HX + X'H').\end{aligned}$$

Consequently, the dual functional can be expressed in terms of X to obtain the functional $\mathbf{J}(X) : \mathcal{X}_+ \to \mathbb{R}$ defined by (23), where $\mathcal{X}_+$ is a convex set.

To prove that the functional (23) has a unique minimizer in $\mathcal{X}_+$ we could now appeal to the proof in [1] that the dual problem in Section VI has a unique solution. However, since now the redundancy in the dual problem has been removed, we can offer a more straightforward alternative proof. We denote by $\bar{\mathcal{X}}_+$ the closure of $\mathcal{X}_+$.

Lemma 6: Suppose that Σ belongs to the range of Γ. Then any nonempty sublevel set

$$\{X \in \bar{\mathcal{X}}_+ \mid \mathbf{J}(X) \le r\} \tag{39}$$

is bounded.

Proof: Let $X \in \bar{\mathcal{X}}_+$ be arbitrary, and define $\lambda := \|X\|$. We want to show that X cannot remain in the level set (39) as $\lambda \to \infty$. To this end, it is no restriction to assume that $\lambda \ge \lambda_0 > 0$. Next set $\tilde{X} := \lambda^{-1} X$ and $\tilde{Q}_\lambda := (\lambda\Psi)^{-1} + G_0^* \tilde{X} + \tilde{X}' G_0$. Then

$$\mathbf{J}(X) = \gamma\lambda - \log\lambda - \mathrm{tr} \int \log \tilde{Q}_\lambda$$

where $\gamma := \mathrm{tr}(H\tilde{X} + \tilde{X}'H')$, and where $\tilde{Q}_\lambda$ depends on λ but is bounded for $\lambda \ge \lambda_0$. First suppose $\gamma > 0$. Then comparing linear and logarithmic growth, $\mathbf{J}(X) \to \infty$ as $\lambda \to \infty$, which contradicts $\mathbf{J}(X) \le r$. Next, suppose that $\gamma \le 0$. Then $\mathbf{J}(X) \to -\infty$ as $\lambda \to \infty$, which contradicts Lemma 5, since $\mathbf{J}(X) = \mathbf{J}(L(\Lambda)B) = \mathbb{J}(\Lambda)$, where $L(\Lambda)$ is the unique solution of the Lyapunov equation (37). Hence, the sublevel set (39) is bounded as claimed.

Lemma 7: The functional $\mathbf{J} : \bar{\mathcal{X}}_+ \to \mathbb{R} \cup \{\infty\}$ has a unique minimizer $\hat{X}$ in $\mathcal{X}_+$.

Proof: We first prove that $\mathbf{J}$, which is continuous on $\mathcal{X}_+$, can be extended as a lower semicontinuous function $\mathbf{J} : \bar{\mathcal{X}}_+ \to \mathbb{R} \cup \{\infty\}$. To this end, let (X_k) be a sequence converging to X in L_∞ norm, and let (Q_k) and Q be the corresponding functions (24), which are continuous on the compact interval $[-\pi, \pi]$, and hence uniformly continuous. Consequently, there is a bound κ such that, for $\theta \in [-\pi, \pi]$, $Q(e^{i\theta}) \le \kappa$, and $Q_k(e^{i\theta}) \le \kappa$ for all k, and hence, by Fatou's lemma

$$-\int \log\left(\frac{Q}{\kappa}\right) \le \liminf_{k\to\infty} -\int \log\left(\frac{Q_k}{\kappa}\right)$$

since $Q_k \to Q$ pointwise. Consequently, $\mathbf{J}(X) \le \liminf_{k\to\infty} \mathbf{J}(X))$, which shows that that $\mathbf{J}$, extended to the boundary $\bar{\mathcal{X}}_+$, is lower semicontinuous. Therefore, it follows from Lemma 6 that the sublevel set (39) is closed and hence bounded. Consequently, by Weierstrass' Theorem, $\mathbf{J}$ has a minimum $\hat{X}$ in $\mathcal{X}$, which must be unique by strict convexity.

It remains to prove that $\hat{X}$ is not the boundary $\partial\mathcal{X}$. To this end, following [16], [18], consider the directional derivative

$$\begin{aligned}\delta\mathbf{J}(X, \delta X) &= \mathrm{tr}\left\{(H\delta X + \delta X' H') \right. \\ &\qquad \left. - \int Q^{-1}\left(G_0^* \delta X + \delta X' G_0\right)\right\} \\ &= \mathrm{tr}\left\{(H\delta X + \delta X'H') - \int Q^{-1}\delta Q\right\}.\end{aligned}$$

Now, for any $X \in \mathcal{X}_+$ and $\bar{X} \in \partial\mathcal{X}$, take $\delta X = X - \bar{X}$ and $X_\lambda = \bar{X} + \lambda\delta X$ and, correspondently, form $\delta Q = Q(z) - \bar{Q}(z)$ and $Q_\lambda(z) = \bar{Q}(z) + \lambda\delta Q_\lambda(z)$, where $\det \bar{Q}(e^{i\theta_0})$ for some $\theta_0 \in [-\pi, \pi]$. Then

$$\delta\mathbf{J}(X_\lambda, -\delta X) = -\mathrm{tr}(H\delta X + \delta X'H') + \int f_\lambda$$

where f_λ is the scalar function

$$f_\lambda(e^{i\theta}) = \mathrm{tr}\{Q_\lambda(e^{i\theta})^{-1}\delta Q(e^{i\theta})\}.$$

Taking the derivative with respect to λ we have

$$\frac{d}{d\lambda} f_\lambda(e^{i\theta}) = \mathrm{tr}\{\delta Q(e^{i\theta})^* Q_\lambda(e^{i\theta})^{-2}\delta Q(e^{i\theta})\} \ge 0$$

and consequently $f_\lambda(e^{i\theta})$ is a monotonically nondecreasing function of λ for all $\theta \in [-\pi, \pi]$. Therefore, as $\lambda \to 0$, f_λ tends pointwise to

$$\begin{aligned}f_0 &= \mathrm{tr}\{\bar{Q}^{-1}(Q - \bar{Q})\} = \mathrm{tr}\{\bar{Q}^{-1}Q - I\} \\ &= \mathrm{tr}\{\bar{Q}^{-1}Q\} - n.\end{aligned}$$

If $\int f_\lambda$ would tend to a finite value as $\lambda \to 0$, (f_λ) would be a Cauchy sequence in $L^1(-\pi, \pi)$ and hence have a limit in $L^1(-\pi, \pi)$ equal almost everywhere to f_0. However, since there is a $\delta > 0$ such that $Q(e^{i\theta}) > \delta$

$$\int f_0 \ge \delta \int \mathrm{tr}\left(\bar{Q}^{-1}\right) - n$$

which is infinite by Proposition 10. Consequently

$$\delta\mathbf{J}(X_\lambda, \bar{X} - X) \to \infty \quad \text{as } \lambda \to 0$$

so there could be no minimum in $\bar{X}$. This concludes the proof.

Since the unique minimizer $\hat{X}$ belongs to the interior $\mathcal{X}_+$, the gradient (27a) is zero there. This proves (26). Then, by (35), the optimal solution of Problem 1 is given by (25). This concludes the proof of Theorem 3.

VIII. Closed-Form Solution for a Special Case of Prior

Next we prove Theorem 4, and hence we now consider the special case where the prior power spectral density is of the particular form

$$\Psi = (G(z)^* \Lambda_0 G(z))^{-1}.$$

Then the matrix function Q defined by (34b) is given by

$$Q(z) = G(z)^*(\Lambda_0 + \Lambda)G(z)$$

which must be positive on the unit circle and hence, by Lemma 12 there exists a constant matrix C such that

$$Q(z) = G(z)^*CC'G(z). \tag{40}$$

We first change the dual functional (34a) by adding the constant $\mathrm{tr}(\Lambda_0\Sigma)$, and compute

$$\begin{aligned}\mathrm{tr}((\Lambda+\Lambda_0)\Sigma) &= \mathrm{tr}\int(\Lambda+\Lambda_0)G\Phi G^*\\ &= \mathrm{tr}\int Q\Phi\\ &= \mathrm{tr}\int C'G\Phi G^*C\\ &= \mathrm{tr}C'\Sigma C\end{aligned} \tag{41}$$

where Φ satisfies (5). In view of (40), the modified functional becomes

$$\begin{aligned}\tilde{\mathbf{J}}(C) &:= \mathbb{J}(\Lambda) + \mathrm{tr}\Lambda_0\Sigma\\ &= \mathrm{tr}\left(C'\Sigma C - \int \log G^*CC'G\right)\end{aligned} \tag{42}$$

which is now a function of C. Recall the following result from Wiener-Masani-Helson-Lowdenslager.

Proposition 8: If $F(z)$ is a square outer matrix-valued function, then

$$\int \log\det FF^* = \log\det F(0)F(0)^*.$$

Proof: The result follows by Jensen's formula after noting that $f = \det F$ is outer ([68, p. 184]).

We now consider once again the functional $\tilde{\mathbf{J}}(C)$ and determine stationarity conditions that provide a form of the optimal C. First

$$\begin{aligned}\tilde{\mathbf{J}}(C) &= \mathrm{tr}\,(C'\Sigma C) - \log\det(B'CC'B)\\ &= \mathrm{tr}\big(C'\Sigma C - \log(B'CC'B)\big).\end{aligned}$$

The gradient with respect to C is

$$\frac{\partial\tilde{\mathbf{J}}}{\partial C} = 2C'\Sigma - 2(B'C)^{-1}B'$$

and hence the stationary point is given by $C'\Sigma = (B'C)^{-1}B'$. This yield the equation

$$B'CC' = B'\Sigma^{-1} \tag{43}$$

for the optimal C, and we readily see that

$$C = \Sigma^{-1}B(B'\Sigma^{-1}B)^{-1/2}$$

satisfies (43). Thus, the optimal Q is

$$\hat{Q}(z) = G(z)^*\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}G(z)$$

and therefore

$$\hat{\Phi}(z) = (G(z)^*\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}G(z))^{-1}.$$

This concludes the proof of Theorem 4.

IX. Conclusion

The topic of the paper is to construct power spectral densities that are consistent with specified moments and are closest to a prior in a suitable sense. The spirit of the work is similar to a long line of contributions going back to [8], including a series of papers [16]–[34] where the emphasis was in identifying and parametrizing power spectra of minimal complexity (i.e., dimensionality of modeling filters). A key tool in these earlier works was a choice of entropy functional that allowed parametrizing solutions via selection of a suitable prior power spectrum. The moment constraints were cast in the form of the state covariance of an input-to-state filter.

In departure from this early work, Ferrante *et al.* [1] proposed to use the KL-divergence between Gaussian probability laws—a formulation that is quite natural from a probabilistic standpoint. The KL-divergence between Gaussian probability laws coincides with the Itakura Saito distance between their respective power spectral densities, and thus, the problem turns out to be equivalent to one studied by Enqvist and Karlsson [2] in the context of scalar processes. The purpose of this paper is to present a simplified alternative optimization procedure that is based on a detailed analysis of the geometry of input-to-state filters and related moment problems. Indeed, the power spectral densities are now parametrized more conveniently by a nonredundant coefficient matrix (X in Theorem 3) containing minimal number of parameters that are necessary. Sections III and IV as well as the proofs later in the paper contain the main contributions.

Appendix

A. Behavior of $\mathbf{J}$ on the Boundary

Lemma 9: Let $\theta \mapsto M(e^{i\theta})$ be a matrix-valued function with Lipschitz-continuous components, and suppose that $M(e^{i\theta})$ is positive semidefinite for all θ and identically zero for $\theta = \theta_0$. Then

$$\int_{-\pi}^{\pi}\mathrm{tr}\{M^{-1}(e^{i\theta})\}\frac{d\theta}{2\pi} = \infty$$

where M^{-1} is defined to have infinite value on any subset of $[-\pi,\pi]$ where it is identically zero.

Proof: Without loss of generality we can assume that $M(e^{i\theta}) = 0$ in an isolated point θ_0. By assumption, we can choose a common Lipschitz constant K and an $\varepsilon > 0$ such that the components $m_{k\ell}(e^{i\theta})$ of M have the bounds

$$\big|m_{k\ell}(e^{i\theta})\big| \le K|\theta-\theta_0|$$

for $|\theta-\theta_0| < \varepsilon$. If $N(e^{i\theta}) := M^{-1}(e^{i\theta})$, its components satisfy

$$\sum_\ell m_{k\ell}(e^{i\theta})n_{\ell k}(e^{i\theta}) = 1 \quad \text{for all } \theta \text{ and } k$$

which then implies that

$$\left|\sum_{\ell} n_{\ell k}(e^{i\theta})\right| K|\theta-\theta_0| \geq 1$$

$$\text{for all } \theta \in (\theta_0-\varepsilon, \theta_0+\varepsilon) \text{ and } k.$$

Consequently, since $N(e^{i\theta}) \geq 0$, there must be a k and an $L>0$ such that

$$n_{kk}(e^{i\theta}) \geq \frac{1}{L|\theta-\theta_0|}, \quad \text{for all } \theta \in (\theta_0-\varepsilon, \theta_0+\varepsilon)$$

and therefore

$$\int_{-\pi}^{\pi} \text{tr}\{M^{-1}(e^{i\theta})\}\frac{d\theta}{2\pi} \geq \frac{1}{L}\int_{\theta_0-\varepsilon}^{\theta_0+\varepsilon} \frac{1}{|\theta-\theta_0|}\frac{d\theta}{2\pi} = \infty$$

as claimed.

Proposition 10: Let $\theta \mapsto Q(e^{i\theta})$ be a matrix-valued function with Lipschitz-continuous components, and suppose that $Q(e^{i\theta})$ is positive semidefinite for all θ and singular for $\theta=\theta_0$. Then

$$\int_{-\pi}^{\pi} \text{tr}\{Q^{-1}(e^{i\theta})\}\frac{d\theta}{2\pi} = \infty.$$

Proof: After applying a constant unitary transformation we can write Q on the form

$$Q = \begin{bmatrix} Q_1 & Q_2 \\ Q_2^* & Q_3 \end{bmatrix}$$

where $Q_1(e^{i\theta_0}) = Q_2(e^{i\theta_0}) = 0$ and $Q_3(e^{i\theta_0}) > 0$. Then

$$Q^{-1} = \begin{bmatrix} \left[Q_1 - Q_2Q_3^{-1}Q_2^*\right]^{-1} & * \\ * & * \end{bmatrix}$$

where the Schur complement

$$M := Q_1 - Q_2Q_3^{-1}Q_2^*$$

is positive semidefinite and has Lipschitz-continuous components. Then the statement of the proposition follows from Lemma 9.

B. Coinvariant Subspaces

Let $\mathcal{H}_2^m$ represent *row* vector-valued functions in the Hardy space of square integrable functions on the circle that have an analytic continuation in the interior of the unit disc—a standard notation $\mathcal{H}_2$ or $\mathcal{H}_2(\mathbb{D})$. The forward shift S amounts to multiplication by z. The backward shift is precisely its adjoint

$$S^* : \mathcal{H}_2^\ell \to \mathcal{H}_2^\ell : x(z) \mapsto \mathbf{\Pi}_{\mathcal{H}_2} z^{-1}x(z).$$

Subspaces, which are invariant under S^*, are those that are orthogonal to invariant subspaces of the forward shift S, i.e., of the form

$$\mathcal{K} := \mathcal{H}_2^{1\times m} \ominus \mathcal{H}_2^{1\times m}V(z)$$

with $V(z)$ an inner (matrix-valued) function, and they are often referred to simply as "coinvariant subspaces." The orthogonal projection onto $\mathcal{K}$ is

$$\mathbf{\Pi}_{\mathcal{K}} : \mathcal{H}_2^{1\times m} \to \mathcal{K}$$

$$x(z) \mapsto \left(\mathbf{\Pi}_{(\mathcal{H}_2^{1\times m})^\perp} x(z)V(z)^*\right)V(z).$$

To see this, note that since $V(z)$ is inner, $\mathbf{\Pi}_{\mathcal{K}}$ defined above is idempotant and Hermitian—hence a projection. It is easy to verify that its kernel is precisely $\mathcal{H}_2^{1\times m}V(z)$, and therefore, $\mathbf{\Pi}_{\mathcal{K}}$ is the orthogonal projection onto $\mathcal{K}$ as claimed.

Let $A \in \mathbb{C}^{n\times n}$ with eigenvalues in $\mathbb{D}$, $B \in \mathbb{C}^{n\times m}$ with (A,B) controllable. Without loss in generality we can always normalize (A,B) so that the corresponding controllability Grammian is the identity I; when this is true

$$AA^* + BB^* = I$$

and $[A,\ B]$ can be competed to a unitary matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

It follows that

$$V(z) = D + zC(I-zA)^{-1}B$$

is an inner matrix-valued function, i.e., it is analytic in $\mathbb{D}$ and $VV^* = V^*V = I$, where $V^* := V(z) := V^*(z^{-1})$. Now, consider

$$G(z) := (I-zA)^{-1}B$$

and the coinvariant subspace $\mathcal{K}$ as noted above. The following statement is known (see [63, Proposition 4]).

Proposition 11: The rows of $G(z)$ form a basis for $\mathcal{K}$.

Proof: The proof is again from [63, Proposition 4]. We first claim that any element in $\mathcal{K}$ is of the form

$$v(zI-A^*)^{-1}C^*V(z) \tag{44}$$

where $v \in \mathbb{C}^{1\times n}$. To see this note that

$$\mathbf{\Pi}_{\mathcal{K}} :\ x_0 + x_1z + \ldots \mapsto$$

$$\left[\mathbf{\Pi}_{(\mathcal{H}_2^{1\times m})^\perp}(x_0 + x_1z + \ldots)\right.$$

$$\left.\times\,(D^* + z^{-1}B^*C^* + \ldots)\right]V(z)$$

$$= v(z^{-1}C^* + z^{-2}A^*C^* + \ldots)V(z)$$

where $v = x_0B^* + x_1B^*A^* + \ldots$. Next, it can be shown [63, Eq. (36)] that

$$G(z) = (zI-A^*)^{-1}C^*V(z). \tag{45}$$

In view of (44), the rows of $G(z)$ span $\mathcal{K}$. Finally, if $vG(z)=0$ for some $v \in \mathbb{C}^{1\times n}$, then necessarily $v=0$ because (A,B) is controllable. Hence, the rows of $G(z)$ are linearly independent and form a basis for $\mathcal{K}$ as claimed.

Lemma 12: Let Λ be a Hermitian $n\times n$-matrix such that

$$Q(z) := G(z)^*\Lambda G(z) > 0$$

for $z = e^{i\theta}$, and $\theta \in [0, 2\pi)$. There exists $\Lambda_o = C_o^* C_o$ with $C_o \in \mathbb{C}^{m\times n}$ such that

$$G(z)^* \Lambda G(z) = G(z)^* \Lambda_o G(z)$$

and $C_o G(z)$ is outer (i.e., minimum phase, that is, stable and stably invertible).

Proof: Since $Q(z)$ is Hermitian and positive definite on the unit circle of the complex plane, it can be factored as

$$Q(z) = a(z)^* a(z)$$

with $a(z)$ outer. But $V(z)G(z)^*\Lambda G(z) = V(z)a(z)^*a(z)$ has all its elements in $\mathcal{H}_2$, since already $V(z)G(z)^*$ does. Since $a(z)$ is outer, $V(z)a(z)^*$ is in $\mathcal{H}_2$ as well. Now, note that $G(z)V(z)^*$ is orthogonal to $\mathcal{H}_2$. Therefore, the zeroth term of $G(z)V(z)^*$ vanishes. It follows that $V(z)a(z)^*a(z)$, which has only positive power of z has no 0th term either. Therefore, $V(z)a(z)^*$ has only positive powers of z and no 0th term. So, finally, we conclude that all elements of $a(z)V(z)^*$ are orthogonal to $\mathcal{H}_2$ and, therefore, the rows of $a(z)$ are in $\mathcal{K}$. Thus, there exists a $C \in \mathbb{C}^{m\times n}$ such that

$$a(z) = CG(z).$$

This completes the proof. ■

REFERENCES

[1] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2561–2575, Oct. 2012.

[2] P. Enqvist and J. Karlsson, "Minimal Itakura-Saito distance and covariance interpolation," in *Proc. 47th IEEE Conf. Decision Control*, 2008, pp. 137–142.

[3] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA, USA: Holden-Day Inc., 1964.

[4] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Dover, 1968.

[5] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proc. 6th Int. Congr. Acoust.*, 1968, vol. 17, pp. C17–C20.

[6] M. Basseville, "Divergence measures for statistical data processing? An annotated bibliography," *Signal Process.*, vol. 93, no. 4, pp. 621–633, 2013.

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

[8] J. P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophys.*, vol. 37, no. 2, pp. 375–376, 1972.

[9] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, no. 9, pp. 963–974, Sep. 1982.

[10] L. Ning, X. Jiang, and T. Georgiou, "Geometric methods for structured covariance estimation," in *Proc. IEEE Amer. Control Conf.*, 2012, pp. 1877–1882.

[11] M. Zorzi and A. Ferrante, "On the estimation of structured covariance matrices," *Automatica*, vol. 48, no. 9, pp. 2145–2151, 2012.

[12] H. Landau, "Maximum entropy and the moment problem," *Bull. Amer. Math. Soc.*, vol. 16, no. 1, pp. 47–77, 1987.

[13] T. Georgiou, "Realization of power spectra from partial covariance sequences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 4, pp. 438–449, Apr. 1987.

[14] T. T. Georgiou, "A topological approach to nevanlinna-pick interpolation," *SIAM J. Math. Anal.*, vol. 18, no. 5, pp. 1248–1260, 1987.

[15] 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*. New York, NY, USA: Springer-Verlag, 1997, pp. 83–108.

[16] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM J. Control Optim.*, vol. 37, no. 1, pp. 211–229, 1998.

[17] T. T. Georgiou, "The interpolation problem with a degree constraint," *IEEE Trans. Autom. Control*, vol. 44, no. 3, pp. 631–635, Mar. 1999.

[18] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM Rev.*, vol. 43, no. 4, pp. 645–675, 2001.

[19] C. I. Byrnes, P. Enqvist, and A. Lindquist, "Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach," *SIAM J. Control Optim.*, vol. 41, no. 1, pp. 23–59, 2002.

[20] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for nevanlinna-pick interpolation with degree constraint," *IEEE Trans. Autom. Control*, vol. 46, no. 6, pp. 822–839, Jun. 2001.

[21] C. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3189–3205, Nov. 2000.

[22] T. T. Georgiou and A. Lindquist, "Kullback-leibler approximation of spectral density functions," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[23] C. I. Byrnes and A. Lindquist, "A convex optimization approach to generalized moment problems," in *Control and Modeling of Complex Systems*. New York, NY, USA: Springer-Verlag, 2003, pp. 3–21.

[24] A. Blomqvist, A. Lindquist, and R. Nagamune, "Matrix-valued nevanlinna-pick interpolation with complexity constraint: An optimization approach," *IEEE Trans. Autom. Control*, vol. 48, no. 12, pp. 2172–2190, Dec. 2003.

[25] T. T. Georgiou, "Solution of the general moment problem via a one-parameter imbedding," *IEEE Trans. Autom. Control*, vol. 50, no. 6, pp. 811–826, Jun. 2005.

[26] C. I. Byrnes and A. Lindquist, "The moment problem for rational measures: Convexity in the spirit of krein, " in, *Modern Analysis and Applications*. New York, NY, USA: Springer-Verlag, pp. 157–169, 2009.

[27] T. T. Georgiou, "Relative entropy and the multivariable moment problem," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[28] C. Byrnes, T. Georgiou, A. Lindquist, and A. Megretski, "Generalized interpolation in ??^{?} with a complexity constraint," *Trans. Amer. Math. Soc.*, vol. 358, no. 3, pp. 965–987, 2006.

[29] C. I. Byrnes and A. Lindquist, "The generalized moment problem with complexity constraint," *Integral Equ. Operator Theory*, vol. 56, no. 2, pp. 163–180, 2006.

[30] M. Pavon and A. Ferrante, "On the georgiou-lindquist approach to constrained Kullback-Leibler approximation of spectral densities," *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 639–644, Apr. 2006.

[31] H. I. Nurdin, "New results on the rational covariance extension problem with degree constraint," *Syst. Control Lett.*, vol. 55, no. 7, pp. 530–537, 2006.

[32] C. I. Byrnes and A. Lindquist, "Important moments in systems and control," *SIAM J. Control Optim.*, vol. 47, no. 5, pp. 2458–2469, 2008.

[33] T. T. Georgiou and A. Lindquist, "A convex optimization approach to ARMA modeling," *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 1108–1119, 2008.

[34] J. Karlsson, A. Lindquist, and A. Ringh, "The multidimensional moment problem with complexity constraint," *Integral Equ. Operator Theory*, vol. 84, no. 3, pp. 395–418, 2016.

[35] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger versus kullback–leibler multivariable spectrum approximation," *IEEE Trans. Autom. Control*, vol. 53, no. 4, pp. 954–967, May 2008.

[36] F. Ramponi, A. Ferrante, and M. Pavon, "A globally convergent matricial algorithm for multivariate spectral estimation," *IEEE Trans. Autom. Control*, vol. 54, no. 10, pp. 2376–2388, Oct. 2009.

[37] A. Ferrante, F. Ramponi, and F. Ticozzi, "On the convergence of an efficient algorithm for kullback–leibler approximation of spectral densities," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 506–515, Mar. 2011.

[38] M. Pavon and A. Ferrante, "On the geometry of maximum entropy problems," *SIAM Rev.*, vol. 55, no. 3, pp. 415–439, 2013.

[39] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, vol. 452. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[40] R. M. Gray, *Linear Predictive Coding and the Internet Protocol: A survey of LPC and a History of of Realtime Digital Speech on Packet Networks*. Delft, The Netherlands: Now Publishers Inc, 2010.

[41] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.
[42] R. E. Kalman, "Realization of covariance sequences," in *Toeplitz Centennial*. New York, NY, USA: Springer-Verlag, 1982, pp. 331–342.
[43] T. T. Georgiou, "Partial realization of covariance sequences," Ph.D. dissertation, Univ. Florida, Gainesville, FL, USA, 1983.
[44] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev, "A complete parameterization of all positive rational extensions of a covariance sequence," *IEEE Trans. Autom. Control*, vol. 40, no. 11, pp. 1841–1857, Nov. 1995.
[45] J. Karlsson and P. Enqvist, "Input-to-state covariances for spectral analysis: The biased estimate," in *Proc. Int. Symp. Math. Theory Netw. Syst.*, 2012, pp. 1–8.
[46] K. Mahata and M. Fu, "Modeling continuous-time processes via input-to-state filters," *Automatica*, vol. 42, no. 7, pp. 1073–1084, 2006.
[47] L. Ying, W. Jian, and S. Zhanjie, "Multivariate spectral estimation based on THREE," *J. China Univ. Posts Telecommun.*, vol. 22, no. 4, pp. 26–32, 2015.
[48] J. Shore and R. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inf. Theory*, vol. 27, no. 4, pp. 472–482, Jul. 1981.
[49] R. S. Ellis, "The theory of large deviations: from boltzmann?s 1877 calculation to equilibrium macrostates in 2d turbulence," *Phy. D, Nonlinear Phenom.*, vol. 133, no. 1, pp. 106–136, 1999.
[50] H.-O. Georgii, "Large deviations and maximum entropy principle for interacting random fields on ZD," *Ann. Probab.*, vol. 21, pp. 1845–1875, 1993.
[51] E. T. Jaynes, "Information theory and statistical mechanics," *Phy. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.
[52] E. Jaynes, "The relation of Bayesian and maximum entropy methods," in *Maximum-entropy and Bayesian Methods in Science and Engineering*. New York, NY, USA: Springer-Verlag, 1988, pp. 25–29.
[53] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[54] I. Sanov, "On the probability of large deviations of random variables," *Mat. Sbornik*, vol. 42, pp. 11–44, 1957.
[55] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[56] L. Y. He, Qi Wang, and G. G. Yin, *System Identification Using Regular and Quantized Observations: Applications of Large Deviations Principles*. New York, NY, USA: Springer-Verlag, 2013.
[57] B. Anderson, J. Moore, and R. Hawkes, "Model approximations via prediction error identification," *Automatica*, vol. 14, no. 6, pp. 615–622, 1978.
[58] A. Stoorvogel and J. Van Schuppen, "System identification with information theoretic criteria," *Nato Asi Ser. F Comp. Syst. Sci.*, vol. 153, pp. 289–338, 1996.
[59] A. Lindquist and G. Picci, *Linear Stochastic Systems*. New York, NY, USA: Springer-Verlag, 2015.
[60] J. Li and P. Stoica, *Robust Adaptive Beamforming*, vol. 88. Hoboken, NJ, USA: Wiley, 2005.
[61] J. Li and P. Stoica, *MIMO Radar Signal Processing*. Hoboken, NJ, USA: Wiley, 2009.
[62] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proc. IEEE*, vol. 78, no. 1, pp. 56–93, Jan. 1990.
[63] T. T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1056–1066, Jul. 2002.
[64] T. T. Georgiou, "The carathéodory–fejér–pisarenko decomposition and its multivariable counterpart," *IEEE Trans. Autom. Control*, vol. 52, no. 2, pp. 212–228, Feb. 2007.
[65] A. Lindquist, "A new algorithm for optimal filtering of discrete-time stationary processes," *SIAM J. Control*, vol. 12, no. 4, pp. 736–746, 1974.
[66] C. I. Byrnes and A. Lindquist, "On the partial stochastic realization problem," *IEEE Trans. Autom. Control*, vol. 42, no. 8, pp. 1049–1070, Aug. 1997.
[67] M. S. Takyar and T. T. Georgiou, "Analytic interpolation with a degree constraint for matrix-valued functions," *IEEE Trans. Autom. Control*, vol. 55, no. 5, pp. 1075–1088, May 2010.
[68] L. V. Ahlfors, Complex Analysis: An Introduction to the Theory of Analytic Functions of one Complex Variable. New York, NY, USA: McGraw-Hill, 1953.

Tryphon T. Georgiou (M'79–SM'99–F'00) received the Diploma in mechanical and electrical engineering from the National Technical University of Athens, Athens, Greece, in 1979, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 1983.

He is currently a Chancellor's Professor in the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA. Dr. Georgiou received the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society for the years 1992, 1999, and 2003. He is a fellow of IFAC and a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).

Anders Lindquist (M'77–SM'86–F'89–LF'10) received the Ph.D. degree in optimization and systems theory from the Royal Institute of Technology, Stockholm, Sweden, in 1972, and an honorary doctorate (Doctor Scientiarum Honoris Causa) from Technion (Israel Institute of Technology) in 2010.

He is currently a Chair Professor at Shanghai Jiao Tong University, Shanghai, China and Professor Emeritus at the Royal Institute of Technology (KTH), Stockholm, Sweden. Before that he had a full academic career in the United States, after which he was appointed to the Chair of Optimization and Systems at KTH.

Dr. Lindquist is a member of the Royal Swedish Academy of Engineering Sciences, a Foreign Member of the Chinese Academy of Sciences, a Foreign Member of the Russian Academy of Natural Sciences, an Honorary Member the Hungarian Operations Research Society, a Fellow of SIAM, and a Fellow of IFAC. He received the 2009 Reid Prize in Mathematics from SIAM and the 2003 George S. Axelby Outstanding Paper Award.