# Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parametrization

Tryphon T. Georgiou

*Abstract*—
**Input spectra which are consistent with a given state covariance of a linear filter, correspond to solutions of an analytic interpolation problem. We derive an explicit formula for the power spectrum with maximal entropy, and provide a linear fraction parametrization of all solutions.**

*Keywords:* **Analytic interpolation, covariance realization.**

## I. Introduction

GIVEN a finite-dimensional linear filter which is driven by a multivariable stationary stochastic process, and given the covariance of the state vector, the family of all input spectra which are consistent with such data correspond to solutions of an analytic interpolation problem [12, 13].

More specifically, consider the (discrete-time) state equations

$$x_k \quad = \quad Ax_{k-1} + Bu_k, \text{ for } k \in \mathbb{Z}. \tag{1}$$

As usual $x_k \in \mathbb{C}^n$, $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, $\operatorname{rank}(B) = m$, $(A, B)$ is a controllable pair, and the eigenvalues of $A$ have modulus $< 1$. Now, if $u_k \in \mathbb{C}^m$ $(k = \ldots, -1, 0, 1, \ldots)$ is a zero-mean stationary stochastic process and

$$\Sigma := \mathcal{E}\{x_k x_k^*\}$$

is the state covariance (with $\mathcal{E}$ being the expectation operator and $^*$ denoting "complex-conjugate transpose"), then it is shown in [13] that the following two *equivalent* conditions hold:

$$\operatorname{rank}\begin{bmatrix} \Sigma - A\Sigma A^* & B \\ B^* & 0 \end{bmatrix} = 2m \tag{2}$$

and,

$$\Sigma - A\Sigma A^* = BH + H^*B^* \text{ for some } H \in \mathbb{C}^{m\times n}. \tag{3}$$

It is also shown that if either (2) or (3) holds and if $\Sigma$ is positive semi-definite, then there exists a suitable stationary input process $u_k$ giving rise to $\Sigma$ as the state covariance of (1).

A finite non-negative matrix-valued measure $d\mu(\theta)$ with $\theta \in (-\pi, \pi]$ represents the power spectrum of a stationary vector-valued stochastic process. It also specifies a matrix-valued function

$$F(\lambda) = \int_{-\pi}^{\pi} \left(\frac{1 + \lambda e^{j\ell\theta}}{1 - \lambda e^{j\ell\theta}}\right) \frac{d\mu(\theta)}{2\pi} + jc, \tag{4}$$

with $jc$ an arbitrary skew-Hermitian constant, which is analytic in the open unit disc $\mathbb{D} := \{\lambda \in \mathbb{C} : |\lambda| < 1\}$ and has non-negative definite real part. In general, the class of such *positive-real* functions $F(\lambda)$ will be denoted by $\mathbb{F}$ and the class of bounded non-negative measures by $\mathbb{M}$. In fact these two families are in exact correspondence via (4) (modulo the skew-Hermitian constant,) and the fact that $d\mu(\theta)$ can be recovered by the radial limits of the real part of $F(\lambda)$ (or, the Hermitian part in case $F(\lambda)$ is matrix-valued; cf. [13]):

$$d\mu(\theta) \sim \lim_{r\nearrow 1} \Re(F(re^{j\theta})). \tag{5}$$

Let $d\mu(\theta)$ represent the spectrum of the input to (1) with $A, B, \Sigma, H$ and $F(\lambda)$ as above. Let also $C \in \mathbb{C}^{m\times n}$, $D \in \mathbb{C}^{m\times m}$ be selected so that

$$V(\lambda) := D + \lambda C(I - \lambda A)^{-1}B \tag{6}$$

is inner, i.e., $V(\xi)^*V(\xi) = I$ for all $|\xi| = 1$, where $I$ is the identity matrix of size determined from the context. (Since $V(\lambda)$ is square, $V(\xi)V(\xi)^* = I$ as well.) Then

$$F(\lambda) = H(I - \lambda A)^{-1}B + Q(\lambda)V(\lambda) \tag{7}$$

for a matrix-valued function $Q(\lambda)$ which is analytic in $\mathbb{D}$. Conversely, if $F(\lambda) \in \mathbb{F}$ and satisfies (7), the real part of $F(\lambda)$ gives rise via (5) to a measure which is consistent with the state-covariance $\Sigma$. Equation (7) is akin to the Nehari problem encountered in $\mathcal{H}_\infty$ control theory, but involves interpolation with positive-real functions instead of functions in $\mathcal{H}_\infty(\mathbb{D})$.

The characterization of state covariances via (2-3) and the connection of admissible power spectra to the analytic interpolation problem (7) are the main results in [13]. The necessity of both (2-3) and (7) are algebraic facts. However, the sufficiency requires the solvability of (7) in $\mathbb{F}$ when $\Sigma \geq 0$. This was accomplished in [13] via "one-step-extensions" and thus, admissible spectra were parametrized by an infinite sequence of contractive parameters. The purpose of the present paper is to focus on the case where $\Sigma > 0$ and provide an explicit formula for the "maximum-entropy" solution as well as give an explicit parametrization of all $\mathbb{F}$-solutions to (7) via a suitable linear fractional transformation (LFT) of a "free parameter".

Department of Electrical and Computer Engineering, University of Minnesota, MN 55455
E-mail: tryphon@ece.umn.edu
Supported in part by the AFOSR, NSF, and MSI.

Analytic interpolation has a history of over 100 years. Hence many of ideas we use are not new. In fact, in the process of developing the theory, we will see the analogs of the Szegö-Geronimus orthogonal polynomials and for the parametrization we will use J-inner/outer factorization *a la* Ball-Helton. The derivations and forumulae are expressed directly in terms of the state matrices of the filter and can be readily implemented as a computer algorithm. A by-product of this approach is an alternative derivations of results in [13].

Besides the purely system theoretic interest in the question of characterizing input spectra consistent with state covariance statistics, the theory gives rise to promising new techniques for high resolution spectral analysis. In fact, the framework we describe encompasses most of the so-called modern nonlinear methods of spectral analysis (like maximum-entropy, MUSIC, ESPRIT, Capon, etc.) and adds a new dimension to the design of such algorithms. This new component is the selection of the state filter (1). Judicious choice of the filter yields a substantial improvement in resolution over state of-the-art. This was demonstrated in [12] for the case of scalar input processes while the present work allows dealing with vectorial processes as well. The potential benefits are highlighted with a (scalar) example where we utilize the maximimum entropy solution. The relevance of such techniques to a truly multivariable situation (e.g., in polarimetric synthetic aperture radar) will be the subject of a future publication.

Section II provides mostly notation. Section III starts off with basics of multivariable prediction theory and develops a formula for the maximum entropy spectrum which is consistent with a given state-covariance matrix $\Sigma > 0$. Section IV develops a linear fractional description of solutions to (7), and hence of input spectra consistent with $\Sigma > 0$. The formulae are worked out in detail only for the case of a non-singular $\Sigma$. Section V gives a characterization of state-covariances of continuous-time filters based on filter parameters— i.e., the continuous-time version of the key result in [13]. Section VI introduces a fractal spectrum (i.e., one with detail at all scales) as a testbed for high resolution algorithms and presents how the maximum entropy estimate performs via simulation. Finally, for completeness, we include in an appendix (Section VII) the proof of a technical lemma on $J$-contractive functions.

## II. Notation and simplifying assumptions

Because (1) is finite-dimensional, we will need to deal mostly with rational functions which are then conveniently expressed via state-matrices. For general $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ matrices of compatible size, we use the notation

$$M(\lambda) =: \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right)$$

to display the state matrices of a transfer function $M(\lambda) = \mathcal{D} + \lambda \mathcal{C}(I - \lambda\mathcal{A})^{-1}\mathcal{B}$ while

$$\left[\begin{array}{cc} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{array}\right]$$

is used to denote the corresponding block matrix of the packed space-state data. A function is said to be *outer* if it is analytic in $\mathbb{D}$ and has an analytic right inverse, *inner* if it is analytic in $\mathbb{D}$ and an isometry on the boundary (e.g., $M(e^{j\theta})^* M(e^{j\theta}) = I$). Similarly, *co-outer* and *co-inner* refer to left invertibility and co-isometry, respectively.

Begin with a given pair $(A, B)$ where $A$ has all its eigenvalues in $\mathbb{D}$. A change of coordinates in (1) transforms $A \mapsto TAT^{-1}$, $B \mapsto TB$ and accordingly $\Sigma \mapsto T\Sigma T^*$ where $T$ an invertible matrix as usual. Such a tranformation can always bring $(A, B)$ into a form where

$$AA^* + BB^* = I. \tag{8}$$

Consider matrices $C, D$ which complete $[A,\ B]$ into the unitary matrix

$$U := \left[\begin{array}{cc} A & B \\ C & D \end{array}\right].$$

It is easy to show that $V(\lambda)$ obtained as in (6) is unitary on the unit circle (hence, both inner and co-inner). When the state matrices transform as above, the state covariance undergoes a congruence trasformation $\Sigma \mapsto T\Sigma T^*$. The normalization of the problem data $(A, B, C, D, \Sigma)$ so that (8) holds, will be assumed throughout as it simplifies the algebra. In particular, the fact that $U$ is unitary provides a number of useful identities such as $AC^* + BD^* = 0$, $A^*A + C^*C = I$, etc.

Given the data $(A, B, \Sigma)$, then a spectral measure $d\mu(\theta) \in \mathbb{M}$ is consistent with it iff

$$\Sigma = \int_{-\pi}^{\pi} \left( G(e^{j\theta}) \frac{d\mu(\theta)}{2\pi} G(e^{j\theta})^* \right), \tag{9}$$

where

$$G(\lambda) := (I - \lambda A)^{-1} B. \tag{10}$$

We denote by $\mathbb{M}_\Sigma$ the set of such consistent distributions:

$$\mathbb{M}_\Sigma = \{ d\mu(\theta) \in \mathbb{M} \ : \ \text{ equation (9) holds} \}.$$

For the most part $(A, B)$ are omitted from the subscript. One exception to this is Remark 4 where the notation $\mathbb{M}_{\Sigma, A, B}$ is used instead.

As usual, $\mathcal{H}_2$ denotes the Hardy space of functions on the circle which are square-integrable and have vanishing negative Fourier coefficients. These have analytic continuation in $\mathbb{D}$. When dealing with vectorial functions these will typically be row-valued. In particular,

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

will represent $1 \times m$ vector-valued functions in $\mathcal{H}_2^{1\times m}$ which are orthogonal to $\mathcal{H}_2^{1\times m} V(\lambda)$. Clearly $\mathcal{H}_2^{1\times m} V(\lambda)$, is invariant under multiplication by the shift operator $\lambda$. (The Beurling-Lax theorem asserts that in fact is the form of all possible such invariant subspaces, e.g., see [17, page 11].) It follows that $\mathcal{K}$ is a subspace which is invariant under the adjoint of $\lambda$—which is the left shift followed by projection onto $\mathcal{H}_2^{p\times m}$. (Accordingly, $\mathcal{K}$ exemplifies all possible "co-invariant" subspaces.)

## III. Multivariate prediction theory

Starting with a multivariable stochastic process $u_k$ having spectral distribution $\mu \in \mathbb{M}$, a Gram matricial structure can be defined on the space of $p \times m$ matrix-valued functions on the circle (see [16, pages 353, 361]) via

$$\begin{aligned}\langle a(\lambda), b(\lambda)\rangle_{d\mu} &:= \int_0^{2\pi} b(e^{j\theta})d\mu(\theta)a(e^{j\theta})^* \qquad (11)\\ &= \mathcal{E}\{(\sum_\ell b_\ell u_{k-\ell})(\sum_\ell a_\ell u_{k-\ell})^*\}\end{aligned}$$

where $a_\ell, b_\ell$ are the Laurent coeffients of $a(\lambda)$, $b(\lambda)$, respectively. Evidently there is a natural correspondence

$$\sum_\ell a_\ell \lambda^\ell \mapsto \sum_\ell a_\ell u_{k-\ell}, \qquad (12)$$

between functions on the unit circle and linear functions of $u_k$, under which the respective gram-matricial inner products in these two spaces agree—where the gram-matricial inner product in the latter is defined as in (12).

A suitable norm (or, possibly, pseudo-norm depending on whether $d\mu$ is positive definite or only semi-definite) is

$$|a|_\mu := [\mathrm{trace}\langle a(\lambda), a(\lambda)\rangle_{d\mu}]^{\frac{1}{2}}.$$

It turns out that the space of functions which are bounded in this norm is complete. This is denoted by $\mathcal{L}_{2,d\mu}^{p\times m}[0, 2\pi]$. Accordingly, the subspace of functions with vanishing negative Fourier coefficients is denoted by $\mathcal{H}_{2,d\mu}^{p\times m}[0, 2\pi]$. The integral in (11) was first defined and studied by Rosenberg and Rosanov (see Masani [16, Sections 5, 6]). In the sequel we explore the geometry of $\mathcal{L}_{2,d\mu}[0, 2\pi]$, albeit focusing mostly on certain finite dimensional subspaces. This makes the analysis especially straightforward. For representing rational functions and carying out the necessary algebra we will use a state-space formalism.

### A. Prediction in $\mathcal{K}^m$

In view of (12), any matrix-valued function

$$h(\lambda) = \sum_{\ell=0}^{\infty} h_\ell \lambda^\ell \qquad (13)$$

with vanishing negative Fourier coefficients and

$$h(0) = h_0 = I \qquad (14)$$

corresponds to

$$h(\lambda) \mapsto u_k - \hat{u}_{k|k-1}$$

which can be interpreted as a "one-step-ahead prediction error". The "predictor", which may not be optimal in any particular way, is the respective linear combination of past values of $u_k$:

$$\hat{u}_{k|k-1} := -\sum_{\ell=1}^{\infty} h_\ell u_{k-\ell}.$$

With this in mind we first seek a least-variance one-step-ahead prediction error which is made up of elements in $\mathcal{K}$. (In the "time domain" this will amount to a linear function of the state vector of (1).) Thus, we seek an element in $\mathcal{K}^m$, i.e., a $m \times n$ matrix-valued function with rows in $\mathcal{K}$, hence of the form

$$\Gamma G(\lambda) \text{ with } \Gamma \in \mathbb{C}^{m\times n}.$$

The variance of the prediction error is

$$\langle \Gamma G(\lambda), \Gamma G(\lambda)\rangle_{d\mu} = \Gamma\langle G(\lambda), G(\lambda)\rangle_{d\mu}\Gamma^* = \Gamma\Sigma\Gamma^*,$$

while the constraint (14) becomes

$$\Gamma B = I.$$

We summarize the facts related to the minimizing solution:

*Lemma 1:* Let $B$ full column rank, $\Sigma > 0$, and

$$\Gamma = (B^*\Sigma^{-1}B)^{-1}B^*\Sigma^{-1}. \qquad (15)$$

Then

$$\Omega := \Gamma\Sigma\Gamma^* < \Gamma_1\Sigma\Gamma_1 \qquad (16)$$

for any $\Gamma_1 \in \mathbb{C}^{m\times n}$ which satisfies $\Gamma_1 \neq \Gamma$ and $\Gamma_1 B = I$. Moreover,

$$\Omega = (B^*\Sigma^{-1}B)^{-1}. \qquad (17)$$

*Proof:* Set $\Gamma_1 = \Gamma + X$ where $XB = 0$ and simply observe that

$$\Gamma_1\Sigma\Gamma_1 = \Gamma\Sigma\Gamma^* + X\Sigma X^*.$$

Since $\Sigma > 0$, $\Gamma$ is the minimizing solution. ■

Thus, the "minimal prediction error" is unique and given by

$$\Phi(\lambda) := \Gamma G(\lambda).$$

*Remark 1:* It is interesting to note that, if $V(\lambda) = V_1(\lambda)V_2(\lambda)\dots V_n(\lambda)$ and accordingly,

$$\mathcal{K}_i = \mathcal{H}_2^m \ominus V_1(\lambda)\dots V_i(\lambda)\mathcal{H}_2^m,$$

then the sequence of $\Phi_i(\lambda)$'s are analogs of the Szegö-Geronimus orthogonal polynomials of the first kind (cf. [7]). Recurrence relations which are analogous to the Szegö-Levinson ones of the classical orthogonal polynomials can be obtained for these matrix-valued functions as well. This is a vast topic for a future occasion. □

It is a classical fact that the Szegö-Geronimus orthogonal polynomials have roots inside the unit circle. In fact *all stable polynomials* arise as orthogonal polynomials for a suitable Pick-Toepliz matrix. The Routh-Hurwitz-Schur-Cohn-like tests for stability in one way or another rely on testing the positivity of the corresponding Pick-Toeplitz matrix which in our case is none other than $\Sigma$. The Pick/Toeplitz matrix is key to a Lyapunov-function-based proof of stability. We now follow such steps for the matricial case at hand:

*Proposition 1:* If $\Sigma > 0$ and $\Phi(\lambda)$ as above, then $\Phi(\lambda)$ is invertible in the closed unit disc $\overline{\mathbb{D}}$.

*Proof:* We can verify by direct algebra that

$$\Phi(\lambda)^{-1} = I - \Gamma A\lambda(I - \lambda(A - B\Gamma A))^{-1}B, \qquad (18)$$

and that

$$\Sigma = B\Omega B^* + (A - B\Gamma A)\Sigma(A - B\Gamma A)^*. \quad (19)$$

Since the pair $(A, B)$ is controllable, so is $(A - B\Gamma A, B)$. Since $\Sigma > 0$, we deduce from (19) that $A - B\Gamma A$ has all eigenvalues in $\mathbb{D}$, and hence $\Phi(\lambda)^{-1}$ is analytic in $\overline{\mathbb{D}}$ as claimed. ■

### B. Maximum entropy solution

The determinant of the variance of error of the optimal one-step-ahead predictor is given by a determinental version of the Szegö-Kolmogorov expression:

$$\mathbb{E}(\mu) := \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det(\dot{\mu}(\theta))d\theta\right\}$$

(cf. [16]). In the above formula "$\dot{\mu}(\theta)$ designates the derivative of $\mu(\theta)$ which exists a.e. on $[-\pi, \pi]$. The exponent

$$\mathbb{I}(\mu) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det(\dot{\mu}(\theta))d\theta$$

is in fact the *entropy rate* of the process.

In our setting, $\mu$ is not known except for the fact that it yields a known state-covariance. Thus, all candidate spectral distributions for $u_k$ satisfy

$$\langle G, G\rangle_{d\mu} = \Sigma.$$

Out of all such spectral distributions we seek one which maximizes the entropy integral $\mathbb{I}(\mu)$, or equivalently $\mathbb{E}(\mu)$.

*Theorem 1:* Let $\Sigma > 0$. The positive-definite measure

$$d\mu_o(\theta) := \left(\Phi(e^{j\theta})^{-1}\Omega^{-1}\left(\Phi(e^{j\theta})^{-1}\right)^*\right)d\theta$$

is the unique solution to

$$\arg\max\{\mathbb{I}(\mu) \,:\, d\mu \geq 0 \text{ and } \langle G, G\rangle_{d\mu} = \Sigma\,\}. \quad (20)$$

Further, the maximal value of the functional is $\log(\det(\Omega))$.

*Proof:* We begin by showing that $\langle G, G\rangle_{d\mu_o} = \Sigma$. To this end we first obtain a minimal realization for the product

$$\begin{aligned} G(\lambda)\Phi(e^{j\theta})^{-1} &= (I - \lambda A)^{-1}B \times \\ &\quad \left(I - \Gamma A\lambda(I - \lambda(A - B\Gamma A))^{-1}B\right) \\ &= (I - \lambda(A - B\Gamma A))^{-1}B =: G_o(\lambda). \end{aligned}$$

The above reduction amounts to removing uncontrollable dynamics. It now follows that

$$\langle G, G\rangle_{d\mu_o} = \int_{-\pi}^{\pi} G_o(e^{j\theta})\Omega G_o(e^{j\theta})^* d\theta$$

is the unique solution of the Lyapunov equation (19). This solution is $\Sigma$.

For the second part we need to show that for any other distribution the entropy integral assumes a larger value. This will follow from comparison of the variance of optimal one-step-ahead prediction errors corresponding to such distributions.

Observe that $\Phi^{-1}$ has vanishing negative Fourier coefficients and a value at the origin equal to the identity. Hence, if $h(\lambda)$ is any function with these same property (i.e., equal to $I$ at the origin and vanishing negative Fourier coefficients), this property is also shared by their product $h\Phi^{-1}$. It follows that

$$\begin{aligned} \langle h, h\rangle_{d\mu_o} &= \langle h\Phi^{-1}, h\Phi^{-1}\rangle_{\Omega d\theta} \\ &\geq \Omega \end{aligned}$$

with equality only if $h(\lambda)\Phi^{-1}(\lambda) \equiv I$. But $\Omega = \langle \Phi, \Phi\rangle_{d\mu_o}$. Thus, $\Phi$ is a unique minimal element in $\mathcal{H}^{m\times m}_{2,\mu_o}$. Therefore,

$$\mathbb{I}(\mu_o) = \log(\det(\Omega)).$$

It is easy to see that if $\mu$ is any other spectral distribution for which $\langle G, G\rangle_{d\mu} = \Sigma$,

$$\langle \Phi, \Phi\rangle_{d\mu} = \Gamma\Sigma\Gamma^* = \Omega. \quad (21)$$

Hence, the variance of the optimal prediction error of such a distribution (i.e., for a choice in (21) possibly other than $\Phi$) is $\leq \Omega$. Therefore

$$\mathbb{I}(\mu) \leq \log(\det(\Omega)) = \mathbb{I}(\mu_o).$$

To see that the above holds in fact with a strict inequality when $\mu \neq \mu_o$, consider two cases. First, assume that $d\mu$ has a nontrivial singular part $d\mu_s$, while $d\mu_a$ denotes its absolutely continuous part. Because $\det(\Phi)$ is nowhere zero on the unit circle $\langle \Phi, \Phi\rangle_{d\mu_s} > 0$. Hence,

$$\langle \Phi, \Phi\rangle_{d\mu} > \langle \Phi, \Phi\rangle_{d\mu_a}.$$

The variance of the optimal prediction error for $d\mu$ is identical to the one corresponding to the absolutely continuous part. The latter however is less than or equal to $\langle \Phi, \Phi\rangle_{d\mu_a}$. This proves the claim for the first case. In the case where $d\mu$ is absolutely continuous, it can be factored as $P(e^{j\theta})\Omega_1 d\theta P(e^{j\theta})^*$ where $P$ is outer and $P(0) = I$. Once again,

$$\langle \Phi, \Phi\rangle_{d\mu} \geq \Omega_1,$$

with equality holding only when $\Phi(\lambda)P(\lambda) \equiv I$ and $\Omega = \Omega_1$ (from (21)). Hence the maximum entropy distribution is unique. This completes the proof. ■

*Remark 2:* Using lower case for denoting as in Geronimus [7] the *normalized* orthogonal polynomial of the first kind, we have that

$$\phi(\lambda) = (B'\Sigma^{-1}B)^{-\frac{1}{2}}B'\Sigma^{-1}(I - \lambda A)^{-1}B.$$

The respective orthogonal polynomial of the second kind, denoted by $\psi(\lambda)$, can be computed by seting

$$\phi(\lambda)^{-1}\psi(\lambda) = F_0(\lambda) + Q(\lambda)V(\lambda),$$

and

$$\phi(\lambda)\psi(\lambda)^* + \psi(\lambda)\phi(\lambda)^* = I.$$

It turns out that

$$\psi(\lambda) = (B'\Sigma^{-1}B)^{-\frac{1}{2}}(I - B'\Sigma^{-1}(I - \lambda A)^{-1}H').$$

We note that in contrast to the scalar interpolation problem (Nevanlinna-Pick or, more generally, Sarason, cf. [9, 10]) where both $\psi(\lambda), \phi(\lambda)$ belong to $\mathcal{K}$, here $\phi(\lambda) \in \mathcal{K}^m$ while $\psi(\lambda) \notin \mathcal{K}^m$.

Using the above expressions it follows that the "maximum-entropy" (positive real) interpolant for which

$$d\mu_o(\theta) = \Re\{F_{ME}(e^{j\theta})\},$$

is

$$\begin{aligned} F_{ME}(\lambda) &= \phi(\lambda)^{-1}\psi(\lambda) \\ &= \left(\begin{array}{c|c} A_{ME} & B_{ME} \\ \hline C_{ME} & D_{ME} \end{array}\right), \end{aligned} \tag{22}$$

where

$$\begin{aligned} A_{ME} &= (I - B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1})A \\ B_{ME} &= B(B'\Sigma^{-1}B)^{-1} + (I - B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1})H' \\ C_{ME} &= -(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}A \\ D_{ME} &= (B'\Sigma^{-1}B)^{-1}(I - B'\Sigma^{-1}H'). \end{aligned}$$

□

## IV. Linear fractional description

In this section we give a *linear fractional transformation* (LFT) description of all $\mathbb{F}$-functions which satisfy (7). This is stated as Theorem 2. The boundary limits of their real part define all spectra consistent with the given state covariance $\Sigma$ as before.

Herein $J$ represents a certain specified signature matrix. The section is organized as follows: we first describe the structure of $J$-inner functions, we then derive formulae for $J$-inner-outer factorization of a matrix whose range contains the graphs of interpolants, and we finally establish an LFT parametetrization with coefficients the entries of the aforementioned $J$-inner factor.

### A. J-expansive functions

First we give a description of $J$-inner matrix functions which are also expansive in $\mathbb{D}$.

*Lemma 2:* Let $J$ be a $q \times q$ signature matrix, $\mathcal{S}$ be a $n \times n$ negative semi-definite matrix, and $A \in \mathbb{C}^{n\times n}$, $\mathcal{D} \in \mathbb{C}^{q\times q}$, $\mathcal{B}, \mathcal{C}' \in \mathbb{C}^{n\times q}$, with $A$ having eigenvalues of modulus less than 1. Define

$$\begin{aligned} \mathcal{U} &:= \begin{bmatrix} A & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}, \\ \mathcal{J} &:= \begin{bmatrix} \mathcal{S} & 0 \\ 0 & J \end{bmatrix}, \text{ and} \\ \mathcal{V}(\lambda) &:= \mathcal{D} + \lambda\mathcal{C}(I - \lambda A)^{-1}\mathcal{B}. \end{aligned} \tag{23}$$

If

$$\mathcal{U}^*\mathcal{J}\mathcal{U} = \mathcal{J}, \tag{24}$$

then $\mathcal{V}(\lambda)$ satisfies

$$\mathcal{V}(\zeta)^*J\mathcal{V}(\zeta) \le J \text{ for } |\zeta| \ge 1, \text{ and} \tag{25}$$

$$\mathcal{V}(\zeta)^*J\mathcal{V}(\zeta) = J \text{ for } |\zeta| = 1. \tag{26}$$

*Proof:* This is a well-known lemma which is often presented in a dual form cast for $J$-contractive functions. For related material see [8]. For completeness of the exposition we provide a proof in the appendix given as Section VII. ■

A matrix-valued function which satisfies (25-26) is said to be *J-lossless (or, inner) on the circle and J-expansive on the unit disc.* We will refer to it as $J$-expansive, for short. An important point is that if $\mathcal{V}(\lambda)$ and $\mathcal{U}$ are as in the lemma, then also

$$\mathcal{U}\mathcal{J}^{-1}\mathcal{U}^* = \mathcal{J}^{-1}, \text{ as well as} \tag{27}$$

$$\mathcal{V}(\zeta)J\mathcal{V}(\zeta)^* \ge J \text{ for } |\zeta| \le 1, \text{ and} \tag{28}$$

$$\mathcal{V}(\zeta)J\mathcal{V}(\zeta)^* = J \text{ for } |\zeta| = 1. \tag{29}$$

In the sequel we will need to consider the following two specific signature matrices

$$J_0 := \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \text{ and } J_1 := \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix},$$

and hence we introduce this "subscript notation" to differentiate when necessary. Obviously, they are congruent since $J_0 = TJ_1T^*$ with $T$ the unitary matrix

$$T := \frac{1}{\sqrt{2}}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}. \tag{30}$$

We conclude the section with a standard lemma pointing to a relationship between elements of a $J_0$-expansive functions (see [5]) (which of course, translated accordingly, is inherited by elements of $J_1$-expansive functions).

*Lemma 3:* Let $J = J_o$ and

$$L_i(\lambda) = \begin{bmatrix} a(\lambda) & b(\lambda) \\ c(\lambda) & d(\lambda) \end{bmatrix}$$

be rational and $J_o$-expansive. Then, $a(\lambda)^{-1}b(\lambda)$ and $c(\lambda)a(\lambda)^{-1}$ are both analytic in the closed unit disc and contractive, i.e., $\|a(\xi)^{-1}b(\xi)\| \le 1$ as well as $\|c(\xi)a(\xi)^{-1}\| \le 1$ for $\|\xi\| \le 1$.

*Proof:* Since $L_i(\xi)JL_i(\xi)^* \ge J$ for all $\xi \in \mathbb{D}$,

$$a(\xi)(a(\xi))^* - b(\xi)(b(\xi))^* \ge I.$$

We conclude that $a(\xi)$ is invertible as well as $\|a(\xi)^{-1}b(\xi)\| < 1$ as claimed. The case of $c(\lambda)a(\lambda)^{-1}$ is similar and requires using the fact that $L_i(\xi)^*JL_i(\xi) \ge J$. ■

### B. J-inner/outer factorization

Most of the steps below require straightforward algebra. The only delicate point is in dealing with the case where the state matrix $A$ is singular. In this case, $D$ is singular as well, due to the fact that $U$ is unitary, and certain algebraic identities hold between the block partitions and their pseudo-inverses. In general, $M^\dagger$ denotes the pseudo-inverse of a matrix $M$, and it holds that $MM^\dagger$ is an orthogonal projection. We will need the fact shown next.

*Lemma 4:* Let

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

be a unitary matrix, then

$$A - BD^\dagger C \quad = \quad (A^*)^\dagger \tag{31}$$

*Proof:* The (Moore-Penrose) pseudoinverse of $A^*$ is the unique solution $X$ of the matrix equations [14, page 243]:

$$\begin{aligned} XA^*X &= X \\ A^*XA^* &= A^* \\ (A^*X)^* &= (XA^*)^* \\ (XA^*)^* &= (A^*X)^*. \end{aligned}$$

For $X = A - BD^\dagger C$ we verify that

$$\begin{aligned} XA^*X &= (A - BD^\dagger C)A^*(A - BD^\dagger C) \\ &= (I - BB^* + BD^\dagger DB^*)(A - BD^\dagger C) \\ &= (I - B(I - D^\dagger D)B^*)(A - BD^\dagger C) \\ &= A - BD^\dagger C - B(I - D^\dagger D)B^*A \\ &\quad -B(I - D^\dagger D)B^*BD^\dagger C \\ &= A - BD^\dagger C = X, \end{aligned}$$

because

$$(I - D^\dagger D)B^*A = (I - D^*(D^*)^\dagger)(-D^*C) = 0$$

while $B(I - D^\dagger D)B^*BD^\dagger C$ is equal to

$$\begin{aligned} & B(I - D^\dagger D)(I - D^*D)D^\dagger C \\ = \;& B(I - D^\dagger D)D^\dagger C - B(I - D^*(D^*)^\dagger)D^*DD^\dagger C \\ = \;& 0. \end{aligned}$$

The remaining identities are shown in a similar manner. ■

*Proposition 2:* Let $A, B, H, \Sigma, G(\lambda), V(\lambda)$ with $\Sigma > 0$ as before. Specialize $J := J_1$, and define

$$K(\lambda) \;:=\; \begin{bmatrix} I & HG(\lambda) \\ 0 & V(\lambda) \end{bmatrix},$$

$$\hat{K}_i(\lambda) := \left(\begin{array}{c|cc} A & -B & H^* \\ \hline HA\Sigma^{-1} & I & HB \\ C\Sigma^{-1} & 0 & D \end{array}\right),$$

$$\hat{K}_o(\lambda) := \left(\begin{array}{c|cc} A & \Sigma^{-1}B & (B - \Sigma^{-1}(BHB + H^*))D^\dagger \\ \hline HA & I & 0 \\ C & 0 & I \end{array}\right),$$

and

$$R := \begin{bmatrix} HA\Sigma^{-1}A^*H^* - HB - B^*H^* & HA\Sigma^{-1}C^* - D^* \\ C\Sigma^{-1}A^*H^* - D & C\Sigma^{-1}C^* \end{bmatrix}.$$

Then

$$K(\lambda) = \hat{K}_o(\lambda)\hat{K}_i(\lambda), \tag{32}$$

$$\hat{K}_o(\lambda) \text{ is outer}, \tag{33}$$

$$-R \text{ is congruent to } J, \tag{34}$$

$$|\det(R)| = 1, \tag{35}$$

and for any constant matrix $R_1$ such that $-R = R_1 J R_1^*$,

$$K_i(\lambda) := R_1^{-1}\hat{K}_i(\lambda)$$

is $J$-expansive.

With $R_1, \hat{K}_o$, etc. as above, we denote

$$K_o(\lambda) := \hat{K}_o(\lambda)R_1.$$

Then

$$K(\lambda) = K_o(\lambda)K_i(\lambda))$$

is a factorization of $K(\lambda)$ into an outer times a $J$-expansive factor.

*Proof:* We directly verify that

$$\hat{K}_o(\lambda)\hat{K}_i(\lambda) = \left(\begin{array}{cc|cc} A & 0 & -B & H^* \\ X & A & \Sigma^{-1}B & Y \\ \hline HA\Sigma^{-1} & HA & I & HB \\ C\Sigma^{-1} & C & 0 & D \end{array}\right),$$

where

$$\begin{aligned} X &= \Sigma^{-1}BHA\Sigma^{-1} + BD^\dagger C\Sigma^{-1} \\ &\quad -\Sigma^{-1}(BHB + H^*)D^\dagger C\Sigma^{-1} \\ Y &= BD^\dagger D - \Sigma^{-1}(BHB + H^*)D^\dagger D + \Sigma^{-1}BHB. \end{aligned}$$

Changing coordinates in the state space of the above system via $\mathcal{A} \mapsto \mathcal{T}^{-1}\mathcal{A}\mathcal{T}$, $\mathcal{C} \mapsto \mathcal{C}\mathcal{T}$, $\mathcal{B} \mapsto \mathcal{T}^{-1}\mathcal{B}$, with

$$\mathcal{T} := \begin{bmatrix} I & 0 \\ -\Sigma^{-1} & I \end{bmatrix}$$

isolates the "unobservable modes" of the product $\hat{K}_o(\lambda)\hat{K}_i(\lambda)$ by bringing the state matrices into the following form

$$\hat{K}_o(\lambda)\hat{K}_i(\lambda) = \left(\begin{array}{cc|cc} A & 0 & -B & H^* \\ 0 & A & 0 & B \\ \hline 0 & HA & I & HB \\ 0 & C & 0 & D \end{array}\right), \tag{36}$$

which, after eliminating the unobservable modes, is precisely $K(\lambda)$. The only nontrivial computation needed to verify (36) is to show that

$$X - \Sigma^{-1}A + A\Sigma^{-1} \;=\; 0, \text{ and} \tag{37}$$

$$Y + \Sigma^{-1}H^* \;=\; B. \tag{38}$$

We first show (37). Begin with the identity $\mathcal{A}^\dagger\mathcal{A} = \mathcal{A}^*(\mathcal{A}^*)^\dagger \Rightarrow \mathcal{A} = \mathcal{A}\mathcal{A}^*(\mathcal{A}^*)^\dagger$, let $\mathcal{A} = \Sigma^{-\frac{1}{2}}A\Sigma^{\frac{1}{2}}$, and multiply the resulting expression on both sides by $\Sigma^{\frac{1}{2}}$. It follows that

$$A\Sigma = (A\Sigma A^*)(A^*)^\dagger. \tag{39}$$

Substituting in the right hand side $A\Sigma A^*$ and $(A^*)^\dagger$ from (3) and (31), respectively, we obtain that

$$A\Sigma = (\Sigma - BH - H^*B^*)(A - BD^\dagger C).$$

Re-ordering terms and multiplying on both sides by $\Sigma^{-1}$ gives (37). To show (38) we first observe that

$$\begin{aligned} H^* + BHB - \Sigma B &= H^* + BHB \\ &\quad -(A\Sigma A^* + BH + H^*B^*)B \\ &= H^* - A\Sigma A^*B - H^*B^*B \\ &= H^* + A\Sigma C^*D - H^*(I - D^*D) \\ &= (A\Sigma C^* + H^*D^*)D, \end{aligned} \quad (40)$$

where we have used (3) in the first step and, subsequently, the fact that $U$ is unitary (hence $B^*B = I - D^*D$ and $A^*B = -C^*D$). Thus,

$$(H^* + BHB - \Sigma B)(I - D^\dagger D) = 0, \quad (41)$$

since, from (40), the rows of the first factor are annihilated by $I - D^\dagger D$. Re-ordering (41) gives (38). Hence, (32) has been established.

Next we argue that $\hat{K}_o(\lambda)$ is outer. This fact can be verified directly by computing the state matrix of $\hat{K}_o(\lambda)^{-1}$ from the realization given in the proposition. This is

$$\begin{aligned} A_o &:= A - \left[\begin{array}{cc} \Sigma^{-1}B, & (B - \Sigma^{-1}(BHB + H^*))D^\dagger \end{array}\right] \left[\begin{array}{c} HA \\ C \end{array}\right] \\ &= A - BD^\dagger C - \Sigma^{-1}BHA + \Sigma^{-1}(BHB + H^*)D^\dagger C \\ &= (A^*)^\dagger - \Sigma^{-1}BH(A^*)^\dagger + \Sigma^{-1}H^*D^\dagger C \\ &= (A^*)^\dagger - \Sigma^{-1}(\Sigma - A\Sigma A^* - H^*B^*)(A^*)^\dagger + \Sigma^{-1}H^*D^\dagger C \\ &= \Sigma^{-1}A\Sigma A^*(A^*)^\dagger + \Sigma^{-1}H^*B^*(A^*)^\dagger + \Sigma^{-1}H^*D^\dagger C \\ &= \Sigma^{-1}A\Sigma + \Sigma^{-1}H^*(B^*(A^*)^\dagger + D^\dagger C) \end{aligned}$$

where in the last step we used (39). The second term vanishes because

$$\begin{aligned} B^*(A^*)^\dagger + D^\dagger C &= B^*(A - BD^\dagger C) + D^\dagger C \\ &= -D^*C - (I - D^*D)D^\dagger C + D^\dagger C \\ &= D^*(I - DD^\dagger)C \\ &= D^*(I - D^*(D^*)^\dagger)C = 0. \end{aligned}$$

Hence $A_o = \Sigma^{-1}A\Sigma$ with all its eigenvalues in $\mathbb{D}$. Therefore $\hat{K}_o(\lambda)$ is outer as claimed.

We now show (34). Pack the system matrices of $\hat{K}_i(\lambda)$ into

$$\hat{\mathcal{K}}_i := \left[\begin{array}{ccc} A & -B & H^* \\ HA\Sigma^{-1} & I & HB \\ C\Sigma^{-1} & 0 & D \end{array}\right].$$

It is easy to verify by direct (but lengthy) algebra that

$$\hat{\mathcal{K}}_i \left[\begin{array}{cc} -\Sigma & 0 \\ 0 & J \end{array}\right] \hat{\mathcal{K}}_i^* = \left[\begin{array}{cc} -\Sigma & 0 \\ 0 & -R \end{array}\right]. \quad (42)$$

We now claim that $\hat{\mathcal{K}}_i$ is invertible. To see this note that

$$\begin{aligned} \det(\hat{\mathcal{K}}_i) &= \det\left(\left[\begin{array}{cc} A & H^* \\ C\Sigma^{-1} & D \end{array}\right] \right. \\ &\qquad \left. - \left[\begin{array}{cc} HA\Sigma^{-1} & HB \end{array}\right] \left[\begin{array}{c} -B \\ 0 \end{array}\right]\right) \\ &= \det\left(\left[\begin{array}{cc} A\Sigma + BHA & H^* + BHB \\ C & D \end{array}\right]\right) / \det(\Sigma) \end{aligned}$$

where in the first step the expression is derived as the determinant of the Schur complement of $\hat{\mathcal{K}}_i$ pivoted about the entry $I$. The second identity follows by simply factoring out $\det(\Sigma)$ from the earlier determinental expression. Thus, we only need to verify that

$$\left[\begin{array}{cc} A\Sigma + BHA & H^* + BHB \\ C & D \end{array}\right]$$

is non-singular. To this end we multiply this matrix on the right with the unitary matrix

$$U^* = \left[\begin{array}{cc} A^* & C^* \\ B^* & D^* \end{array}\right]$$

to obtain

$$\left[\begin{array}{cc} \Sigma & A\Sigma C^* + H^*D^* \\ 0 & I \end{array}\right]$$

which is invertible. Thus, $\hat{\mathcal{K}}_i$ is invertible and moreover,

$$\det(\hat{\mathcal{K}}_i) = 1. \quad (43)$$

From (42) it follows that readily that $-R$ is invertible and congruent to $J$, i.e., it has the same signature. From (43) and (42) we also conclude (35).

We finally show that $\mathcal{K}_i$ is $J$-expansive. If $\mathcal{K}_i$ denotes the corresponding matrix containing state-space data, i.e.,

$$\mathcal{K}_i = \left[\begin{array}{cc} I & 0 \\ 0 & R_1^{-1} \end{array}\right] \hat{\mathcal{K}}_i,$$

then

$$\mathcal{K}_i \left[\begin{array}{cc} -\Sigma & 0 \\ 0 & J \end{array}\right] \mathcal{K}_i^* = \left[\begin{array}{cc} -\Sigma & 0 \\ 0 & J \end{array}\right]. \quad (44)$$

Using Lemma 2, the claim is established. ■

*Remark 3:* It is interesting to pursue the algebra leading to (42) a bit further. If $S := -R^{-1}$ then

$$\mathcal{K}_i^* \left[\begin{array}{cc} \Sigma^{-1} & 0 \\ 0 & S \end{array}\right] \mathcal{K}_i = \left[\begin{array}{cc} \Sigma^{-1} & 0 \\ 0 & J \end{array}\right]. \quad (45)$$

By writing explicitely the relations between known quantities and the entries of $S$ it is easy to see that the $(1,1)$-block entry of $S$ is

$$S_{11} = -B^*\Sigma^{-1}B.$$

By virtue of the fact that $S = -R^{-1}$, this relates to the inverse of a corresponding Schur compliment of $R$, in fact:

$$(R_{22} - R_{12}R_{11}^{-1}R_{21})^{-1} = B^*\Sigma^{-1}B.$$

The expression on the left hand side appears in the parametric description of solutions given in [13] via one-step-extensions. Thus, $(B^*\Sigma^{-1}B)^{-1/2}$ represents the "left" uncertainty radius while (already given in [13]) $(C\Sigma^{-1}C^*)^{-1/2}$ represents the "right" uncertainty radius. Throughout, all inverses exist since $B, C^*$ are assumed to have full column rank and $\Sigma$ is positive definite. (If any of these assumptions is omitted then analogous formulae are still valid with pseudo-inverses replacing the inverses. However, the algebra for verifying these is significantly more tedious.) □

*Remark 4:* We explain another way of dealing with a singular state matrix $A$. To this end we temporarily modify the notation $\mathbb{M}_\Sigma$ into $\mathbb{M}_{\Sigma,A,B}$, so that the parameters $A, B$ of the relevant filter are shown explicitely. Let $L \in \mathbb{C}^{m\times n}$ be such that $A-BL$ has eigenvalues in $\mathbb{D}-\{0\}$ and consider the system

$$\begin{aligned} x_k &= (A-BL)x_{k-1} + B(u_k + Lx_{k-1}) && (46)\\ &= A_L x_{k-1} + Bv_k. && (47)\end{aligned}$$

The state covariance is again $\Sigma$ while the spectral measure of the input process $v_k$ is now

$$d\mu_v(\theta) = (I + e^{j\theta}LG(e^{j\theta})d\mu(\theta)(I + e^{j\theta}LG(e^{j\theta})^*. \quad (48)$$

Since

$$(I + \lambda LG(\lambda)^{-1} = I - \lambda L(I - \lambda A_L)^{-1}B$$

is analytic, $d\mu(\theta)$ can be determined from $d\mu_v(\theta)$ using (48) as well. Hence there is a bijective correspondence between the two solution sets

$$d\mu(\theta) \in \mathbb{M}_{\Sigma,A,B} \Leftrightarrow d\mu_v(\theta) \in \mathbb{M}_{\Sigma,A_L,B}$$

and the data of the problem modified so that the "new" state matrix is nonsingular. □

*Remark 5:* Although it is very easy to effect the factorization $-R = R_1 J R_1^*$ (e.g., using singular value decomposition of $R$), we do not know if there is a particular value for the factor $R_1$ which can be expressed directly in terms $A, B, C, D, H$ and $\Sigma$ in "closed form". Such an expression would be highly desirable. □

We continue with the main result of the section which characterizes all positive real solutions to (7) via a suitable LFT.

### C. LFT Parametrization

We denote by $\mathbb{S}_o$ the open unit ball of matrix-valued function in $\mathcal{H}_\infty$, i.e.,

$$\mathbb{S}_o := \left\{ Y(\lambda) \ : \text{analytic in } \mathbb{D} \text{ and } \sup_{\xi\in\mathbb{D}} \|Y(\xi)\| < 1 \right\}.$$

For simplicity of the presentation we restrict our attention to solutions of (7) in the interior $\mathbb{F}$ which will be denoted by

$$\mathbb{F}_o := \left\{ F(\lambda) \ : \text{analytic in } \mathbb{D} \text{ and } \inf_{\xi\in\mathbb{D}} \Re\{F(\xi)\} > 0 \right\}.$$

Here, since $F(\xi)$ is a matrix, $\Re$ signifies the "Hermitian part of". The two sets $\mathbb{S}_o$ and $\mathbb{F}_o$ are in bijective correspondence via

$$F(\lambda) = (I + Y(\lambda))^{-1}(I - Y(\lambda))$$

and the parametrization given below can equally well be expressed in terms of a free parameter belonging to either of these sets.

*Theorem 2:* Let $A, B, H, \Sigma, G(\lambda), V(\lambda)$ with $\Sigma > 0$ as before, and $K_i(\lambda)$ as in Proposition 2. Then, a function $F(\lambda)$ is in $\mathbb{F}_o$ and satisfies

$$F(\lambda) = HG(\lambda) + Q(\lambda)V(\lambda) \quad (49)$$

with $Q(\lambda)$ analytic in $\mathbb{D}$, if and only if

$$F(\lambda) = F_1(\lambda)^{-1}F_2(\lambda) \quad (50)$$

with

$$\begin{bmatrix} F_1(\lambda), & F_2(\lambda) \end{bmatrix} = \begin{bmatrix} I + Y(\lambda), & I - Y(\lambda) \end{bmatrix} K_i(\lambda) \quad (51)$$

and $Y(\lambda) \in \mathbb{S}_o$.

*Proof:* We first show the "if" part. So let

$$F(\lambda) = F_1(\lambda)^{-1}F_2(\lambda)$$

with (51) being valid and $Y(\lambda) \in \mathbb{S}_o$. Then

$$\frac{1}{2}\begin{bmatrix} F_1(\lambda) + F_2(\lambda), & F_1(\lambda) - F_2(\lambda) \end{bmatrix} = \begin{bmatrix} I, & Y(\lambda) \end{bmatrix} L_i(\lambda)$$

where $L_i(\lambda) = TK_i(\lambda)T$ and $T$ as in (30). Clearly, $L_i(\lambda)$ is $J_o$-expansive. With $a(\lambda), c(\lambda)$ as in Lemma 3 we conclude that,

$$\begin{aligned} F_1(\lambda) + F_2(\lambda) &= a(\lambda) + Y(\lambda)c(\lambda)\\ &= a(\lambda)\left(I + Y(\lambda)c(\lambda)a(\lambda)^{-1}\right)\end{aligned}$$

is invertible in the closed unit disc. Since $L_i(\lambda)$ is $J_o$-expansive and $Y(\lambda)$ strictly contractive,

$$\begin{bmatrix} F_1(\xi) + F_2(\xi), & F_1(\xi) - F_2(\xi) \end{bmatrix} J_o \begin{bmatrix} F_1(\xi)^* + F_2(\xi)^* \\ F_1(\xi)^* - F_2(\xi)^* \end{bmatrix}$$

is strictly positive for $\xi \in \mathbb{D}$. From this it easily that $F_1(\xi) + F_2(\xi)$ is invertible for all $\xi$ in the closed unit disc and that

$$(F_1(\xi) + F_2(\xi))^{-1}(F_1(\xi) - F_2(\xi)) \in \mathbb{S}_o.$$

It follows that $F_1(\xi)$ is invertible $\mathbb{D}$ and that $F(\lambda) \in \mathbb{F}_o$. To see that $F(\lambda)$ also satisfies (7) simply observe that

$$\begin{bmatrix} I & F_1(\lambda)^{-1}F_2(\lambda) \end{bmatrix} = \begin{bmatrix} Z_1(\lambda), & Z_2(\lambda) \end{bmatrix} K(\lambda)$$

which follows from (51) and

$$\begin{aligned} \begin{bmatrix} Z_1(\lambda), & Z_2(\lambda) \end{bmatrix} &:= F_1(\lambda)^{-1} \times \\ &\qquad \begin{bmatrix} I + Y(\lambda), & I - Y(\lambda) \end{bmatrix} K_o(\lambda)^{-1}.\end{aligned}$$

Clearly $Z_1(\lambda) = I$ and (7) follows.

We now argue the validity of the converse. First note that the set of solutions to (7), herein denoted by

$$\begin{aligned}\mathcal{P} \;:=\; & \{F(\lambda) = F_o(\lambda) + Q(\lambda)V(\lambda) \in \mathbb{F}_o \\ & \text{with } Q(\lambda) \text{ analytic in } \mathbb{D}\},\end{aligned}$$

is convex. From the "if" part we know that how to construct a particular element $f_0(\lambda) \in \mathcal{P}$ of the required form (50-51). In particular, we take $Y(\lambda) = 0$. Then

$$\left[\begin{array}{cc} F_1(\lambda) & F_2(\lambda) \end{array}\right] = \left[\begin{array}{cc} I, & I \end{array}\right] K_i(\lambda)$$

and

$$\begin{aligned}\left[\begin{array}{cc} I, & f_0(\lambda) \end{array}\right] &:= \left[\begin{array}{cc} I, & F_1(\lambda)^{-1}F_2(\lambda) \end{array}\right] \\ &= \left[\begin{array}{cc} F_1(\lambda)^{-1}, & F_1(\lambda)^{-1} \end{array}\right] K_i(\lambda) \\ &=: \left[\begin{array}{cc} m_0(\lambda), & n_0(\lambda) \end{array}\right] K_i(\lambda)\end{aligned}$$

with $m_0(\lambda)$ invertible on the closed unit disc. Now take any other $f_1(\lambda) \in \mathcal{P}$ and write it in the form

$$\begin{aligned}\left[\begin{array}{cc} I & f_1(\lambda) \end{array}\right] &= \left[\begin{array}{cc} I, & q_1(\lambda) \end{array}\right] K(\lambda) \\ & \left[\begin{array}{cc} m_1(\lambda), & n_1(\lambda) \end{array}\right] K_i(\lambda).\end{aligned}$$

We claim that $m_1(\lambda)$ is invertible in the closed unit disc. If this is not the case, then consider the continuous path

$$f_\tau(\lambda) := (1-\tau)f_0(\lambda) + \tau f_1(\lambda) \in \mathcal{P}$$

with $\tau \in [0,1]$ and denote $m_\tau, n_\tau$ the corresponding convex combination of $m_0(\lambda), m_1(\lambda)$, and $n_0(\lambda), n_1(\lambda)$, respectively. It follows that for some intermediate value $\tau = \tau_0$, $m_{\tau_0}(\lambda)$ is singular on the boundary of $\mathbb{D}$. Since $K_i(\lambda)$ is $J$-unitary on the boundary, it follows that the real part of $f_{\tau_0}(\lambda)$ is singular on the boundary. This contradicts our assumption that $f_\tau(\lambda) \in \mathcal{P}$. Hence, $m_1(\lambda)$ is invertible. It can be also seen that $m_1(\lambda)^{-1}n_1(\lambda) \in \mathbb{F}_o$. We finally write

$$\begin{aligned}\left[\begin{array}{cc} I, & f_1(\lambda) \end{array}\right] &= m_1(\lambda)\left[\begin{array}{cc} I, & m_1(\lambda)^{-1}n_1(\lambda) \end{array}\right] K_i(\lambda) \\ &= m_1(\lambda)(I+Y(\lambda))^{-1} \times \\ &\quad \left[\begin{array}{cc} I+Y(\lambda), & I-Y(\lambda) \end{array}\right] K_i(\lambda)\end{aligned}$$

with $Y(\lambda) \in \mathbb{S}_o$ and

$$m_1(\lambda)^{-1}n_1(\lambda) = (I+Y(\lambda))^{-1}(I-Y(\lambda)).$$

This completes the proof. ■

## V. Continuous-time filters

Analogous results hold in the case of continuous-time processes and systems. We summarize the analog of Theorem 1 for this case along with a short proof.

*Theorem 3:* Let $B \in \mathbb{C}^{n\times m}$, $A \in \mathbb{C}^{n\times n}$ be such that the continuous-time system

$$\dot{x}(t) = Ax(t) + Bu(t),\; t \in \mathbb{R}$$

is a controllable, the eigenvalues of $A$ have negative real part, and let $\Sigma \in \mathbb{C}^{n\times n}$ be a Hermitian positive definite matrix. The matrix $\Sigma$ is the stationary state covariance of the above system for a suitable input process if and only if

$$\operatorname{rank}\left[\begin{array}{cc} \Sigma A + A^*\Sigma & B \\ B^* & 0 \end{array}\right] = \operatorname{rank}\left[\begin{array}{cc} 0 & B \\ B^* & 0 \end{array}\right]. \quad (52)$$

*Proof:* The necessity of the condition follows easily as in the discrete-time case [13]. Simply write

$$\Sigma = \int_{-\infty}^{\infty} (j\omega I - A)^{-1}Bd\mu(\omega)B^*(-j\omega I - A^*)^{-1}$$

where $d\mu(\omega)$ a nonnegative measure on $\mathbb{R}$, and utilize the algebraic identities $A(sI-A)^{-1} = s(sI-A)^{-1} - I$ and $(-sI-A^*)^{-1}A^* = -s(-sI-A^*)^{-1} - I$ to express $A\Sigma + \Sigma A^*$ in the form

$$A\Sigma + \Sigma A^* = -BH - H^*B^*, \quad (53)$$

with

$$H = \int_{-\infty}^{\infty} \left(d\mu(\omega)B^*(-j\omega I - A^*)^{-1}\right).$$

The rank condition claimed is equivalent to the solvability of (53), e.g., see [13, Proposition 1].

For the sufficiency we need to produce a suitable spectrum that generates $\Sigma$. To this end consider a rational power spectrum of the form

$$d\mu_o(\omega) := \Phi(j\omega)^{-1}\Omega\frac{d\omega}{2\pi}\left(\Phi(j\omega)^{-1}\right)^*,$$

with

$$\begin{aligned}\Phi(s) &:= I + C_1(sI-A)^{-1}B, \\ C_1 &:= \frac{1}{2}(B^*\Sigma^{-1}B)^{-1}B^*\Sigma^{-1} - H\Sigma^{-1}, \text{ and} \\ \Omega &:= (B^*\Sigma^{-1}B)^{-1}.\end{aligned}$$

We compute that

$$\begin{aligned}\Phi(s)^{-1} &= I - C_1(sI-A_1)^{-1}B \text{ and} && (54) \\ G(s)\Phi(s)^{-1} &= (sI-A_1)^{-1}B, && (55)\end{aligned}$$

where $A_1 = A - BC_1$. From (53) (which is equivalent to the rank condition in the statement), it follows that

$$A_1\Sigma + \Sigma A_1^* = -B\Omega B^*. \quad (56)$$

Note that $(A_1, B\Omega B^*)$ is a controllable pair since $(A,B)$ is, and also that $B\Omega B^* \geq 0$. From Lyapunov theory we know that the above are sufficient to conclude that $A_1$ has all its eigenvalues in the left half of the complex plane, and that the solution $\Sigma$ of (56) is also given by

$$\begin{aligned}\Sigma &= \int_{-\infty}^{\infty} (j\omega I - A_1)^{-1}B\Omega\frac{d\omega}{2\pi}B^*(-j\omega - A_1^*)^{-1}. \\ &= \int_{-\infty}^{\infty} (j\omega I - A)^{-1}Bd\mu_o(\omega)B^*(-j\omega I - A^*)^{-1}.\end{aligned}$$

This completes the proof. ■

*Remark 6:* It is important to point out that, in general, output covariances of linear systems can be characterized neither by a rank condition, nor in terms of the solvability of only a linear algebraic equation—cf. [11, Remark 2, page 786]. □

## VI. High resolution analysis: An example

The following academic example is intended to highlight the potential relevance of the theory in high resolution spectral analysis. We make no effort to compare with other techniques except for the periodogram.

We consider a rather nontraditional spectrum with "detail at all scales"—which is appropriate as a benchmark for spectral analysis techniques. The ideal power spectrum is "self similar". The spectral distribution function, shown in Figure 1, is the so-called "devil's staircase" and is constructed as follows. Begin with an interval $[f_1, f_2]$ (in our case $[0, \pi]$) where we set $\mu(f_1) = 0$ and $\mu(f_2) = 1$. Define distribution $\mu(\theta)$ to be constant and equal to $\frac{\mu(f_2)-\mu(f_1)}{2}$ over the middle third of the interval

$$[f_1 + \frac{f_2 - f_1}{3}, f_1 + 2\frac{f_2 - f_1}{3}],$$

and repeat the above indefinitely for the remaining two thirds $[f_1, f_1 + \frac{f_2-f_1}{3}]$ and $[f_1 + 2\frac{f_2-f_1}{3}, f_2]$. The spectral distribution we construct in this way is continuous from 0 to 1 with derivative equal to 0 everywhere except on a set of zero measure where it is not defined. The derivative of $\mu$ which is the "spectral density" can be thought of as consisting of infinitely many spectral lines of "zero amplitude" each. It represents the power spectrum of a deterministic signal which is impossible to draw. An approximation of it by a finite number of sinusoids of nonzero amplitude–corresponding to the above process for drawing $\mu$ terminated after finitely many steps–is drawn in Figure 2. It examplifies the fact that *more lumps of spectral energy become apparent at finer and finer scales.*

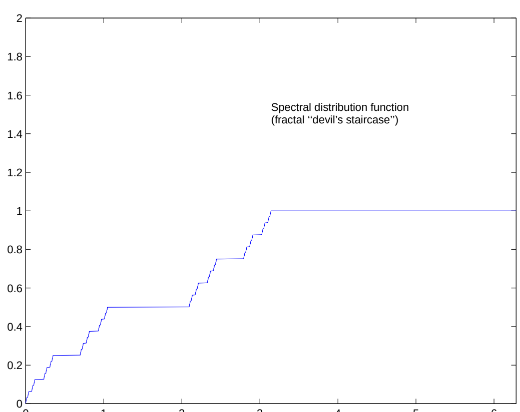


Fig. 1.

An approximation of a process with such a spectrum was generated by combining $2^9 = 1024$ exponential sinusoids,

$$u_k = \sum_{\ell=1}^{2^9} \frac{1}{2^9} e^{j(\theta_\ell k + \phi_\ell)},$$

with $\phi_\ell$ random and uniformly distributed in $[0, 2\pi]$. A typical realization for $u_k$ is shown in Figure 3.

Figure 4 shows a periodogram which was constructed based on 1024 data points. Subplots (3,2) and (3,3) zoom

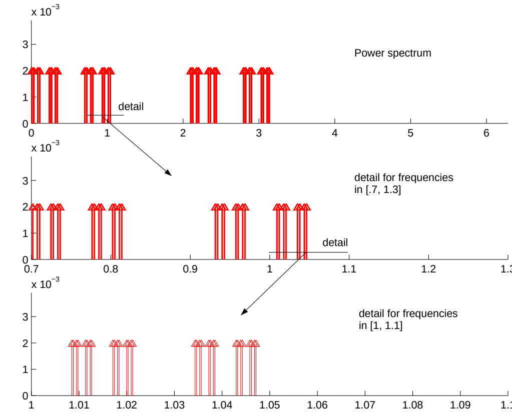


Fig. 2.

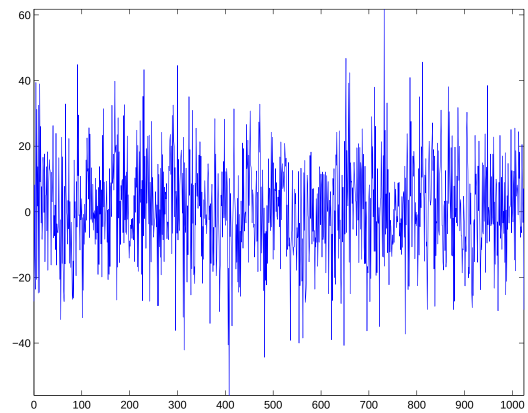

Fig. 3.

in onto the intervals $[0.7, 1.3]$ and $[1, 1.1]$, respectively. It is apparent that the resolution of the peridogram does not go beyond the two spectral lumps which are located around frequencies 1.01 and 1.04, respectively.

We present three sets of spectra constructed using the maximum entropy formalism and estimated state-covariances. The input-to-state filters that were utilized in each case have *progressively narrower bandpass characteristic centered around* 1.02. These filters are designed as follows. Consider a state matrix $A_o$ consisting of a single Jordan block

$$\begin{bmatrix} \lambda_0 & 1 & 0 & \dots \\ 0 & \lambda_0 & 1 & \vdots \\ \dots & & & \dots \\ 0 & \dots & & 1 \end{bmatrix}$$

with eigenvalue $\lambda_0 = \rho e^{j\theta_0}$ and let $B_o = [0,\ 0, \dots 0,\ 1]'$. Then, the "pass-band" of the corresponding input-to-state frequency response $\|G(e^{j\theta})\|$ is centered around $\theta_0$. Typically, $\theta_0$ is selected in the frequency range of interest. For $\rho = 0$, $\theta_0$ is irrelevant and the filter consists of simple delay elements with $\|G(e^{j\theta})\|$ constant over frequencies. In this case, the state covariance is simply an ordinary Toeplitz matrix made up of the covariance lags of $u_k$ and the maximum-entropy formula in Theorem 2 reduces to the

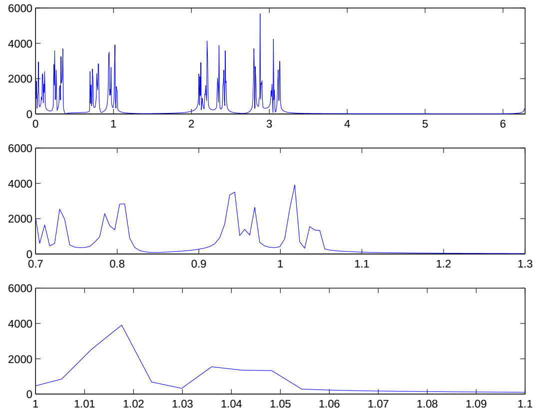


Fig. 4.

classical one (e.g., see [19]). In general, we have observed that "balancing" the state matrices via similarity transformation $A_o \mapsto A = TA_oT^{-1}$ and $B_o \mapsto B = TB_o$ so that $AA^* + BB^* = I$ is beneficial for numerical reasons. This is followed in our simulations.

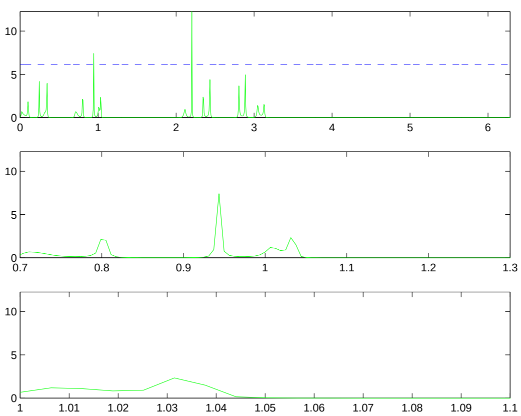


Fig. 5.

For the three spectra in Figures 5-7 we took ($n = 65, \lambda_0 = 0$), ($n = 20, \lambda_0 = 0.65e^{j1.02}$), and ($n = 7, \lambda_0 = 0.95e^{j1.01}$), respectively. In each case, we give three subplots which differ as to the range of values for the frequency axis—zooming in on to the interval of interest. The first subplot contains in addition the frequency response $\|G(e^{j\theta}\|$ of the corresponding filter drawn with a dotted curve and suitably normalized.

The spectrum shown in Figure 5 amounts to an ordinary maximum entropy based on 65 samples of the autocorrelation function. The resolution in the neighborhood of, e.g., $[1, 1.1]$ is comparable to that of the periodogram.

In the spectrum in Figure 6 we observe that the spectral lumps about 1.015 and 1.045 are more clearly pronounced. Yet the next finer set of spectral lumps is not discernible.

Moving on to the spectrum shown in Figure 7, we observe that the spectral lump at 1.015 has now been resolved into two adjacent lumps (centered about 1.01 and 1.02 respectively) at the expense of accuracy/resolution elsewhere

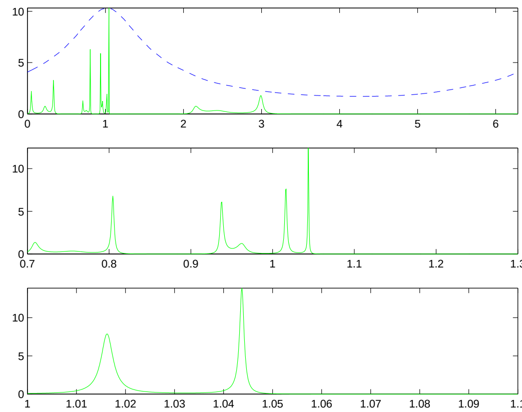


Fig. 6.

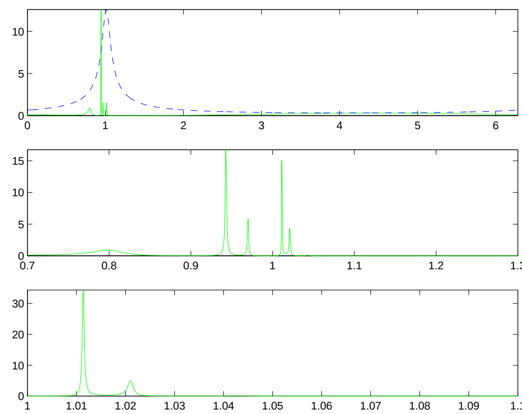


Fig. 7.

on the frequency axis.

We make the following further observations:

(i) As the pass band of the filters becomes progressively narrower, their reliability diminishes outside the bandpass, while inside, the resolution increases. This may be traced to the fact that, in view of (9), the state covariance depends more heavily on the part of the spectrum within the passband.

(ii) Values of $\rho$ close to 1, adversely affect variability of the estimated state covariances. However, at the same time a smaller size $n$ for the filter and the corresponding state-covariance can be afforded.

(iii) The limitations of methods based on second-order statistics are due to statistical and numerical errors in estimating the state-covariances. Thus, they are of a different nature than the time-bandwidth limitations of periodogram-based methods, and thus can exceed the resolution of Fourier transform methods under suitable conditions.

The example we presented (i.e., one of a fractal spectrum) is suitable as a benchmark for testing high resolution methods. The particular exercise in applying the maximum entropy method while computing the state covariance via a suitably tuned input-to-state filters, suggests the potential of the approach—though this is far from conclusive.

Additional studies giving guidelines for filter design, and establishing the limits to performance and the variance of the estimators are needed.

## VII. Appendix: Proof of Lemma 2

Consider the dynamical system

$$\begin{aligned} x_k &= \zeta A x_{k-1} + \mathcal{B} u_k, \\ y_k &= \zeta C x_{k-1} + \mathcal{D} u_k, \end{aligned}$$

where $x_{-1} = 0$, $k = 0, 1, \ldots$, and $\zeta \in \mathbb{D}$. With

$$\mathcal{U}_\zeta := \begin{bmatrix} \zeta A & \mathcal{B} \\ \zeta \mathcal{C} & \mathcal{D} \end{bmatrix},$$

using equation (24) and the fact that $\mathcal{S}$ is negative semi-definite, we deduce that

$$\begin{aligned} \mathcal{U}_\zeta^* \mathcal{J} \mathcal{U}_\zeta &= \begin{bmatrix} \bar\zeta A^* & \bar\zeta \mathcal{C}^* \\ \mathcal{B}^* & \mathcal{D}^* \end{bmatrix} \begin{bmatrix} \mathcal{S} & 0 \\ 0 & J \end{bmatrix} \begin{bmatrix} \zeta A & \mathcal{B} \\ \zeta \mathcal{C} & \mathcal{D} \end{bmatrix} \\ &= \begin{bmatrix} |\zeta|^2 \mathcal{S} & 0 \\ 0 & J \end{bmatrix} \geq \mathcal{J}, \end{aligned}$$

then that

$$x_k^* \mathcal{S} x_k - y_k^* J y_k \geq x_{k-1}^* \mathcal{S} x_{k-1} - u_k^* J u_k,$$

for all $k \geq 0$, and then

$$\sum_{k=0}^{\ell} (u_k^* J u_k - y_k^* J y_k) = x_{\ell+1}^* \mathcal{S} x_{\ell+1} \leq 0 \tag{57}$$

for all $\ell \geq 0$. I.e., $x_{\ell+1}^* \mathcal{S} x_{\ell+1}$ represents a "storage" function. If we define the $\ell \times \ell$-block matrices

$$J_\ell := \begin{bmatrix} J & & & \\ & J & & \\ & & \ddots & \\ & & & J \end{bmatrix}, \text{ and}$$

$$\mathcal{T}_{\zeta,\ell} := \begin{bmatrix} \mathcal{D} & & & \\ \zeta \mathcal{C}\mathcal{B} & \mathcal{D} & & \\ \vdots & & & \\ \zeta^\ell \mathcal{C} A^{\ell-1} \mathcal{B} & \ldots & \zeta \mathcal{C}\mathcal{B} & \mathcal{D} \end{bmatrix},$$

(which are block diagonal, and block lower triangular, respectively), then from (57) follows that

$$\mathcal{T}_{\zeta,\ell}^* J_\ell \mathcal{T}_{\zeta,\ell} \geq J_\ell. \tag{58}$$

Since the eigenvalues of $\zeta A$ are inside the open unit disc, the series

$$\mathcal{V}_\ell(\zeta) := \mathcal{D} + \zeta \mathcal{C}\mathcal{B} + \zeta^2 \mathcal{C} A \mathcal{B} + \ldots + \zeta^\ell \mathcal{C} A^{\ell-1} \mathcal{B}$$

converges as $\ell \to \infty$, while (58) implies that

$$\mathcal{V}_\ell(\zeta)^* J \mathcal{V}_\ell(\zeta) \geq J$$

for all $\ell$. Tt follows that

$$\mathcal{V}(\zeta)^* J \mathcal{V}(\zeta) \geq J \text{ for all } \zeta \in \mathbb{D}. \tag{59}$$

It remains to show that (59) holds with equality on the unit circle. This is a consequence of the algebraic identity

$$V^*(\lambda^{-1}) J V(\lambda) = J$$

which is shown using (24):

$$\begin{aligned} V^*(\lambda^{-1}) J V(\lambda) &= \mathcal{D}^* J \mathcal{D} \\ &\quad + \lambda^{-1} \mathcal{B}^* (I - \lambda^{-1} A^*)^{-1} \mathcal{C}^* J \mathcal{D} \\ &\quad + \lambda \mathcal{D}^* J \mathcal{C} (I - \lambda A)^{-1} \mathcal{B} \\ &\quad + \mathcal{B}^* (I - \lambda^{-1} A^*)^{-1} \mathcal{C}^* J \mathcal{C} (I - \lambda A)^{-1} \mathcal{B} \\ &= J - C^* \mathcal{S} C \\ &\quad - \lambda^{-1} \mathcal{B}^* A^* (I - \lambda A^*)^{-1} \mathcal{S} \mathcal{B} \\ &\quad - \lambda \mathcal{B}^* \mathcal{S} (I - \lambda A)^{-1} A \mathcal{B} \\ &\quad + \mathcal{B}^* (I - \lambda^{-1} A^*)^{-1} (\mathcal{S} - A^* \mathcal{S} A)(I - \lambda A)^{-1} \mathcal{B} \\ &= J. \end{aligned}$$

## References


[1] J. A. Ball and J.W. Helton, A Beurling-Lax theorem for the Lie group $U(m,n)$ which contains most classical interpolation theory," *J. Op. Theory,* **9:** 107-142, 1983.

[2] S. Yuliar, M.R. James, and J.W. Helton, Dissipative control systems synthesis with full state feedback, preprint, November 1995.

[3] C. I. Byrnes, T.T. Georgiou, and A. Lindquist, A new approach to spectral estimation: A tunable high-resolution spectral estimator, *IEEE Trans. on Signal Proc.*, **48(11):** 3189-3206, November 2000.

[4] P. Delsarte, Y. Genin, and Y. Kamp, Schur parameterization of positive definite block-Toeplitz systems, *SIAM J. Appl. Math.*, **36(1):** 34-46, 1979.

[5] P. Dewilde and H. Dym, Lossless chain scattering matrices and optimum linear predition: the vector case, *Circuit theory and applications*, **9:** 135-175, 1981.

[6] B.A. Francis, **A course in $H_\infty$ control theory**, Springer-Verlag, 1987.

[7] Ya. L. Geronimus, **Orthogonal Polynomials**, English translation from Russian by Consultants Bureau, New York, 570 pages, 1961.

[8] Y. Genin, P. Van Dooren, T. Kailath, J-M. Delosme, and M. Morf, On $\Sigma$-lossless transfer functions and related questions, *Linear Algebra and Applications*, **50:** 251-275, 1983.

[9] T.T. Georgiou, "A topological approach to Nevanlinna-Pick interpolation", *SIAM J. on Mathematical Analysis*, 18(5): 1248-1260, September 1987.

[10] T.T. Georgiou, "The interpolation problem with a degree constraint," *IEEE Trans. on Automatic Control*, **44(3)**: 631-635, March 1999.

[11] T.T. Georgiou, Signal Estimation via Selective Harmonic Amplification: MUSIC, Redux, *IEEE Trans. on Signal Processing*, March 2000, **48(3)**: 780-790.

[12] T.T. Georgiou, Spectral Estimation via Selective Harmonic Amplification, *IEEE Trans. on Automatic Contr.*, January 2001, **46(1)**: 29-42.

[13] T.T. Georgiou, The state covariance and its relation to the power spectrum of the input, *IEEE Trans. on Automatic Contr.*, submitted for publication, preprint, January 2001.

[14] G. H. Golub and C. F. Van Loan, **Matrix Computations**, The Johns Hopkins University Press, Baltimore, 1989.

[15] S. Haykin, **Nonlinear Methods of Spectral Analysis,** Springer-Verlag, New York, 247 pages, 1979.

[16] P. Masani, Recent trends in multivariate prediction theory, in **Multivariate Analysis**, P.R. Krishnaiah, Ed., Academic Press, pp. 351-382, 1966.

[17] M. Rosenblum and J. Rovnyak, **Hardy Classes and Operator Theory**, Oxford University Press, 1985.

[18] V.P. Potapov, The multiplicative structure of $J$-contractive matrix functions, *Amer. Math. Soc. Transl.*, **15**: 131-243, 1960.

[19] P. Stoica and R. Moses, **Introduction to Spectral Analysis**, Prentice Hall, 1997.