The Carathéodory–Fejér–Pisarenko Decomposition and Its Multivariable Counterpart

Tryphon T. Georgiou, *Fellow, IEEE*

***Abstract*—When a covariance matrix with a Toeplitz structure is written as the sum of a singular one and a positive scalar multiple of the identity, the singular summand corresponds to the covariance of a purely deterministic component of a time-series whereas the identity corresponds to white noise—this is the Carathéodory–Fejér–Pisarenko (CFP) decomposition. In the present paper we study multivariable analogs for block-Toeplitz matrices as well as for matrices with the structure of state-covariances of finite-dimensional linear systems (which include block-Toeplitz ones). To this end, we develop theory which addresses questions of existence, uniqueness and realization of multivariable power spectra, possibly having deterministic components. We characterize state-covariances which admit only a deterministic input power spectrum, and we explain how to realize multivariable power spectra which are consistent with singular state covariances via decomposing the contribution of the singular part. We then show that multivariable decomposition of a state-covariance in accordance with a "deterministic component + white noise" hypothesis for the input does not exist in general. We finally reinterpret the CFP-dictum and consider replacing the "scalar multiple of the identity" by a covariance of maximal trace which is admissible as a summand. The summand can be either (block-)diagonal corresponding to white noise or have a "short-range correlation structure" correponding to a moving average component. The trace represents the maximal variance/energy that can be accounted for by a process at the input (e.g., noise) with the aforementioned structure, and this maximal solution can be computed via convex optimization. The decomposition of covariances and spectra according to the range of their time-domain correlations is an alternative to the CFP-dictum with potentially great practical significance.**

***Index Terms*—Pisarenko harmonic decomposition, short-range correlation, spectral analysis.**

I. Introduction

PRESENT day signal processing is firmly rooted in the analysis and interpretation of second order statistics. In particular, the observation that singularities in covariance matrices reveal a deterministic linear dependence between observed quantities, forms the basis of a wide range of techniques, from Gauss' least squares to modern subspace methods in time-series analysis. In the present work we study the nature and origin of singularities in certain structured covariance matrices which arise in multivariable time-series.

Historically, modern subspace methods (e.g., MUSIC, ESPRIT) can be traced to Pisarenko's harmonic decomposition and even earlier to a theorem by Carathéodory and Fejér on a canonical decomposition of finite Toeplitz matrices [26], [27], [32]. The Toeplitz structure characterizes covariances of stationary scalar time-series. In this context, singularity of the Toeplitz matrix reflects the fact that there is a *unique* power spectrum which is consistent with the data. Their multivariable counterpart, block-Toeplitz matrices, having a less stringent structure, has received considerably less attention. The present work was motivated by the apparent lack of analogues of the Carathéodory–Fejér–Pisarenko (CFP) decomposition for finite block-Toeplitz matrices as well as for the more general setting of state-covariances of a known linear dynamical system. In particular, the singularity of a block-Toeplitz covariance matrix does not reflect in general a unique consistent power spectrum.

The decomposition of a covariance matrix into non-negative definite summands suggests a corresponding additive decomposition of the relevant time-series. Such a decomposition may provide insight and facilitate modeling and stochastic realization of the underlying random process—for analysis or filtering purposes. While our declared objective has been to study multivariable generalizations of the CFP decomposition, a substantial part of the paper is devoted to results that pertain to the existence, uniqueness and realization of power spectra, possibly with deterministic components.

The realization of power spectra from second order statistics amounts to a moment problem, and as such, in certain important cases, it can be cast in the language of analytic interpolation theory. This indeed is the case for the type of problems studied here. Consequently, we make extensive use of techniques and insights gained in recent years in the analytic interpolation literature, as presented in e.g., [1]–[3], [6], [12], [15], [16], [25], [29], and [31]. Following our interest in deterministic power spectra we develop certain explicit conditions which guarantee uniqueness. These appear to be new. We also study the so-called "central solution in the singular case," because this relates to deterministic components in the power spectrum.

After we develop basic results on singular state-covariances and associated power spectra we return to the CFP-dictum and show that, in general, *it does not hold for multivariable processes*. Indeed, we show that the decomposition of a given state-covariance in accordance with a "deterministic component + white noise" hypothesis for the input does not exist in general. We are then led to reinterpret the CFP-dictum. We seek a maximal-trace summand of the given state-covariance which can be interpreted as due to input noise (white noise or a moving average process of any given order). The trace represents the maximal variance/energy that can be accounted for, and the optimal

Manuscript received September 12, 2005. Recommended by Associate Editor H. Hjalmarsson. This work was supported in part by the National Science Foundation and the Air Force Office of Scientific Research.

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

Digital Object Identifier 10.1109/TAC.2006.890479

0018-9286/$25.00 © 2007 IEEE

solution can be computed via convex optimization. This reinterpretation allows the decomposition of covariances and spectra according to the range of time-domain correlations for the corresponding processes. It is an alternative to the CFP-dictum with potentially great practical relevance.

Briefly, the contents of this paper are as follows. In Section II, we begin with background material on matrices with the structure of a state-covariance of a known linear dynamical system—block-Toeplitz matrices being a special case. Section III discusses the connection between covariance realization and Carathéodory (analytic) interpolation. Section IV presents a duality between left and right matricial Carathéodory interpolation and their relation to the time arrow in dynamical systems generating the state-process. Duality is taken up again in Section V where we study optimal prediction and postdiction (i.e., prediction backwards in time) of a stochastic input based on state-covariance statistics. The variance of optimal prediction and postdiction errors coincide with left and right uncertainty radii in a Schur representation of the family of consistent spectra given in [19], [20] and elucidate the symmetry observed in these references. Further, Section V presents geometric conditions on the state-covariance for the input process to be deterministic and for the optimal predictor and postdictor to be uniquely defined. Vanishing of the variance of the optimal prediction or postdiction errors is shown in Section VI to characterize state-covariances for which the family of consistent input spectra is a singleton.

Section VII gives a closed form expression for the power spectrum corresponding the "central solution" of [20] (also, e.g., [16, Ch. V]). This expression is valid in the case where the state-covariance is singular. Naturally, the subject of this section has strong connections with the theory of Szegö–Geronimus orthogonal polynomials and their multivariable counterparts [8], [12, Ch. 8], which now become matricial functions sharing the eigen-structure of the transfer function of the underlying dynamical system. Then, Section VIII explains how to isolate the deterministic component of the power spectrum via computation of relevant residues with matrix techniques.

Having completed the analysis of singular state-covariances and their relation to input power spectra, we show in Section IX that a state-covariance may not admit a decomposition into one corresponding to white-noise plus another corresponding to a deterministic input. To this end, a natural interpretation of the CFP dictum is to seek a maximal white-noise component at the input consistent with a given state-covariance. We explain how this is computed and discuss a further generalization where the input "noise" is allowed to have "short-range correlation structure." In particular, if the state-covariance is $\ell \times \ell$ (block-)Toeplitz, then we may seek to account for input noise whose auto-covariance vanishes after the first k moments, i.e., colored noise modeled by at most a k-order moving average filter. In this way, a maximal amount of variance that may be due to short range correlations can be accounted for, leaving the remaining energy/variance to be attributed to periodic deterministic components and possibly, stochastic components with long range (longer than k) correlations. Further elaboration of this viewpoint, for the special case of Toeplitz matrices, is given in the follow-up paper [22].

II. Structured Covariance Matrices

Throughout we consider a multivariable, discrete-time, zero-mean, stochastic process $\{u_k : k \in \mathbb{Z}\}$ taking values in $\mathbb{C}^{m\times 1}$ with $m \in \mathbb{N}$. Thus, u_k is to be thought of as a column vector. We denote by $R_k := \mathcal{E}\{u_\ell u_{\ell-k}^*\}$, for $k, \ell \in \mathbb{Z}$, the sequence of matrix covariances and by $d\mu(\theta)$ the corresponding matricial spectral measure for which

$$R_k = \frac{1}{2\pi}\int_0^{2\pi} e^{-jk\theta} d\mu(\theta)$$

for $k \in \mathbb{Z}$ (see, e.g., [30]). As usual, star $(^*)$ denotes the complex-conjugate transpose, prime $(')$ denotes the transpose, $j := \sqrt{-1}$ following the usual "engineering" convention, and $\mathcal{E}\{\,\cdot\,\}$ denotes the expectation operator. Whenever star $(^*)$ is applied to a rational function of z it represents the paraconjugate Hermitian $f(z)^* := f^*(z^{-1})$ where $f^*(\,\cdot\,)$ refers to $*$-ing the coefficients of $f(\,\cdot\,)$ whereas the transformation of the argument is indicated separately.

It is well-known that a covariance sequence $\{R_\ell : \ell \in \mathbb{Z}$ and $R_{-\ell} = R_\ell^*\}$ is completely characterized by the non-negativity of the block-Toeplitz matrices

$$\mathbf{R}_\ell := \begin{bmatrix} R_0 & R_1 & \dots & R_\ell \\ R_{-1} & R_0 & \dots & R_{\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ R_{-\ell} & R_{-\ell+1} & \dots & R_0 \end{bmatrix} \tag{1}$$

for all ℓ. That is, such an infinite sequence with the property that $\mathbf{R}_\ell \geq 0, \forall \ell$, qualifies as a covariance sequence of a stochastic process and vice versa. On the other hand, the infinite sequence of R_ℓ's defines the spectral measure $d\mu$ (up to an additive constant) and conversely.

It is often the case that only a finite set of second-order statistics is available, and then, it is of interest to characterize possible extensions of the finite covariance sequence $\{R_0, R_1, \dots, R_\ell\}$, or equivalently, the totality of consistent spectral measures (see [8]–[11], [4], [19], and [20]). In general, these are no longer specified uniquely by the finite sequence $\{R_0, R_1, \dots, R_\ell\}$. In the present paper we are interested in particular, in the case where a finite set of second-order statistics such as $\{R_0, R_1, \dots, R_\ell\}$ completely specifies the corresponding spectral measure (and, hence, any possible infinite extension as well). We address this question in the more general setting of structured covariance matrices which includes block-Toeplitz matrices as a special case.

A block-Toeplitz matrix such as $\mathbf{R}_\ell$ given in (1) can be thought of as the state-covariance of the linear (discrete-time) dynamical system

$$x_k = Ax_{k-1} + Bu_k, \qquad \text{for } k \in \mathbb{Z}. \tag{2}$$

where

$$A = \begin{bmatrix} O_m & O_m & \dots & O_m & O_m \\ I_m & O_m & \dots & O_m & O_m \\ & \ddots & \ddots & \vdots & \vdots \\ & & & & \\ O_m & O_m & & I_m & O_m \end{bmatrix} \quad B = \begin{bmatrix} I_m \\ O_m \\ \vdots \\ O_m \end{bmatrix} \tag{3}$$

with O_m and I_m the zero and the identity matrices of size $m \times m$, A $(\ell+1) \times (\ell+1)$ and B $(\ell+1) \times 1$ block matrices, respectively. The size of each block is $m \times m$ and hence the actual sizes of A, B are $n \times n$ and $n \times m$, with $n = (\ell+1)m$, respectively. While for general state-matrices A, B the structure of the state-covariance may not be visually recognizable, it is advantageous, for both, economy of notation and generality, to develop the theory in such a general setting—the theory of block-Toeplitz matrices being a special case.

Thus, henceforth, we consider an *input-to-state* dynamical system as in (2) where

$$\begin{aligned}
&\text{(4a)} \quad u_k \in \mathbb{C}^m, \quad x_k \in \mathbb{C}^n, \quad A \in \mathbb{C}^{n\times n}, \quad B \in \mathbb{C}^{n\times m}\\
&\text{(4b)} \quad \text{rank}(B) = m\\
&\text{(4c)} \quad (A, B) \text{ is a reachable pair, and}\\
&\text{(4d)} \quad \text{all the eigenvalues of } A\\
&\qquad\quad \text{have modulus} < 1. \qquad (4)
\end{aligned}$$

Without loss of generality and for convenience we often assume that the pair (A, B) has been normalized as well so that

$$\text{(4e)} \quad AA^* + BB^* = I_n.$$

Conditions (4a-d) are standing assumptions throughout. Whenever condition (4e) is assumed valid, this will be stated explicitly. With $u_k \in \mathbb{C}^m, k \in \mathbb{Z}$, a zero-mean stationary stochastic process, we denote by

$$\mathbf{R} := \mathcal{E}\{x_k x_k^*\}$$

the corresponding (stationary) state-covariance. The space of Hermitian $n \times n$ matrices will be denoted by $\mathbb{H}_n \subset \mathbb{C}^{n\times n}$ while positive (resp. nonegative) definiteness of an $\mathbf{R} \in \mathbb{H}_n$ will be denoted by $\mathbf{R} > 0$ (respectively, $\mathbf{R} \geq 0$). Any state-covariance as before certainly satisfies both conditions, i.e., it is Hermitian and non-negative definite. The following statement characterizes the linear structure imposed by (2).

Theorem 1: (see [19, Thms. 1 and 2]): Let 4(a)–(d) hold. A nonnegative–definite Hermitian matrix $\mathbf{R}$ (i.e., $\mathbb{H}_n \ni \mathbf{R} \geq 0$) arises as the (stationary) state-covariance of (2) for a suitable stationary input process $\{u_k\}$ if and only if the following equivalent conditions hold:

$$\begin{aligned}
&\text{(5a)} \quad \text{rank}\begin{bmatrix} \mathbf{R} - A\mathbf{R}A^* & B \\ B^* & 0 \end{bmatrix} = 2m\\
&\text{(5b)} \quad \mathbf{R} - A\mathbf{R}A^* = BH + H^*B^*, \quad H \in \mathbb{C}^{m\times n}. \quad (5)
\end{aligned}$$

A finite $m \times m$ non-negative matrix-valued measure $d\mu(\theta)$ with $\theta \in [0, 2\pi)$ represents the power spectrum of a stationary $m \times 1$-vector-valued stochastic process. The class of all such $m \times m$ matrix-valued non-negative bounded measures will be denoted by $\mathbb{M}$. Note that the size m is suppressed in the notation because it will be the same throughout. Starting with a stationary input u_k with power spectral distribution $d\mu \in \mathbb{M}$, the state-covariance of (2) can be expressed in the form of the integral (cf. [30, Ch. 6])

$$\mathbf{R} = \int_0^{2\pi} \left(G(e^{j\theta}) \frac{d\mu(\theta)}{2\pi} G(e^{j\theta})^* \right) \qquad (6)$$

where

$$G(z) := (I_n - zA)^{-1}B \qquad (7)$$

is the transfer function of (2) (with z corresponding to the delay operator, so that "stability" corresponds to "analyticity in the open unit disc $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$"). Thus, either condition (5a) or (5b) in the above theorem characterizes the range of the mapping

$$\mathbb{M} \ni d\mu \mapsto \mathbf{R}$$

specified by (6). The family of power spectral distributions which satisfy (6) will be denoted by

$$\mathbb{M}_{\mathbf{R}} := \{d\mu(\theta) \in \mathbb{M}: \text{ equation (6) holds}\}.$$

The previous theorem states that this family is nonempty when $\mathbf{R}$ satisfies the stated conditions. Furthermore, a complete parametrization of $\mathbb{M}_{\mathbf{R}}$ is given in [19], [20].

The present work explores the case where $\mathbb{M}_{\mathbf{R}}$ is a singleton. The special case where u_k is scalar and $\mathbf{R}$ a Toeplitz matrix (but not "block-Toeplitz") goes back to the work of Carathéodory and Fejér a century ago, and later on, to the work of Pisarenko (see [26], [27], [32]). In the scalar case, $\mathbb{M}_{\mathbf{R}}$ is a singleton if and only if $\mathbf{R}$ is singular (and of course non-negative definite). Then u_k is deterministic with a spectral distribution $d\mu$ having at most $n-1$ discontinuities (spectral lines). In the present paper we obtain analogous results when $\mathbf{R}$ is a state-covariance and $\mathbb{M}_{\mathbf{R}}$ is a singleton, and then we study decomposition of a general $\mathbf{R} > 0$ into a covariance due to "noise" plus a singular covariance with deterministic components—in the spirit of the CFP decomposition of Toeplitz covariance matrices.

III. Connection With Analytic Interpolation

The early work of Carathéodory and Fejér was motivated by questions in analysis which led to the development of analytic interpolation theory—a subject which has since attained an important place in operator theory, and more recently, closer to home, in robust control engineering. We review certain rudimentary facts and establish notation.

A non-negative measure $d\mu \in \mathbb{M}$ specifies an $m \times m$ matrix-valued function

$$\begin{aligned}
F(z) &= \int_0^{2\pi} \left(\frac{1 + ze^{-j\theta}}{1 - ze^{-j\theta}} \right) \frac{d\mu(\theta)}{2\pi} + jc\\
&=: \mathcal{H}[d\mu] + jc \qquad (8)
\end{aligned}$$

with jc an arbitrary skew-Hermitian constant (i.e., $c \in \mathbb{H}_m$), which is analytic in the open unit disc $\mathbb{D}$ and has non-negative definite Hermitian part (see, e.g., [10, p. 36]). We denote by $\mathcal{H}[d\mu]$ the *Herglotz* integral given in the previous line. The class of such $m \times m$ functions with non-negative Hermitian part in $\mathbb{D}$, herein denoted by

$$\mathbb{F} := \{F(z) : F(z) = \mathcal{H}[d\mu] + jc, c \in \mathbb{H}_m d\mu \in \mathbb{M}\}$$

is named after Carathéodory and often referred to simply as "positive-real." Conversely, given $F \in \mathbb{F}$, a corresponding $d\mu(\theta) \in \mathbb{M}$ can be recovered from the radial limits of the Hermitian part of $F(z)$. It is obtained via

$$d\mu(\theta) = \lim_{r \nearrow 1} \mathrm{Herm}\{F(re^{j\theta})\}d\theta$$

interpreted as a weak limit or, alternatively via

$$\mu(\theta) = \lim_{r \nearrow 1} \int_0^\theta \mathrm{Herm}\{F(re^{j\tau})\}d\tau + c_0 \tag{9}$$

for $c_0 \in \mathbb{H}_m$ (cf. [8, Sec. II]). In fact the families $\mathbb{F}$ and $\mathbb{M}$ are in correspondence via (8) and (9) (with elements in $\mathbb{F}$ identified if they differ by a skew-Hermitian constant and distributions defined up to an additive constant).

With (4e) in place, select $C \in \mathbb{C}^{m \times n}, D \in \mathbb{C}^{m \times m}$ so that

$$U := \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{10}$$

is a unitary matrix. Then

$$V(z) := D + zC(I_n - zA)^{-1}B \tag{11}$$

is inner, i.e., $V(\xi)^*V(\xi) = V(\xi)V(\xi)^* = I_m$ for all $|\xi| = 1$. The rows of $G(z)$ form a basis of

$$\mathcal{K} := \mathcal{H}_2^{1\times m} \ominus \mathcal{H}_2^{1\times m}V(z) \tag{12}$$

where $\mathcal{H}_2$ denotes the Hardy space of functions analytic in $\mathbb{D}$ with square-integrable boundary limits. This can be easily seen from the identity ([19], (38))

$$G(z) = (zI - A^*)^{-1}C^*V(z) \tag{13}$$

(from which it follows that the entries of $G(z)V(z)^*$ are in $\mathcal{H}_2^\perp$, the orthogonal complement of $\mathcal{H}_2$ in the Lebesgue space of square-integrable function on the unit circle $\mathcal{L}_2(\partial\mathbb{D})$).

Now let $d\mu(\theta)$ represent the power spectrum of the input to (2), $\mathbf{R}$ the corresponding state-covariance, and $F(z)$ obtained via (8). Then, $\mathbf{R}$ turns out to be the Hermitian part of the operator

$$\mathcal{W} : \mathcal{K} \to \mathcal{K} : \nu(z) \mapsto \mathbf{\Pi}_\mathcal{K}(\nu(z)F(z)^*) \tag{14}$$

with respect to basis elements being the rows of $G(z)$, where $\mathbf{\Pi}_\mathcal{K}$ denotes the orthogonal projection onto $\mathcal{K}$ (see [19, eqs. (40) and (41)]). Of course, $\mathbf{R}$ is also the Grammian $\langle G(z), G(z)\rangle_{d\mu}$ with respect to the inner product

$$\langle g_i(z), g_k(z)\rangle_{d\mu} := \int_0^{2\pi} \left(g_i(e^{j\theta})\frac{d\mu(\theta)}{2\pi}g_k(e^{j\theta})^*\right).$$

This is in fact the content of (6).

The relationship between $F(z)$ and $\mathbf{R}$ can be obtained by way of $\mathcal{W}$. If H is the zeroth Fourier coefficient of $G(z)F(z)^*$ then the matrix representation W for $\mathcal{W}$ with respect to the rows of $G(z)$ satisfies (see [19])

$$W - AWA^* = H^*B^* \tag{15}$$

leading to (5) for $\mathbf{R} = W + W^*$. The matrices W or H completely specify $\mathbf{\Pi}_\mathcal{K}F(z)^*|_\mathcal{K}$ and in fact

$$F(z) = F_0(z) + Q(z)V(z) \tag{16}$$

with $F_0(z) := H(I_n - zA)^{-1}B$ and $Q(z)$ is a matrix-valued function which is analytic in $\mathbb{D}$. Conversely, if $F(z) \in \mathbb{F}$ and satisfies (16), then it gives rise via (9) to a measure which is consistent with the state-covariance $\mathbf{R}$.

Equation (16) specifies a problem akin to the Nehari problem in $\mathcal{H}_\infty$-control theory involving positive-real functions instead (cf. [11], [12], [1]–[3], and [15]).

IV. A Dual Formalism

Using (13), (6) can be rewritten as

$$\mathbf{R} = \int_0^{2\pi} \left(G_r(e^{j\theta})\frac{d\mu_r(\theta)}{2\pi}G_r(e^{j\theta})^*\right) \tag{17}$$

where $G_r(z) = (zI - A^*)^{-1}C^*$ and

$$d\mu_r(\theta) = V(e^{j\theta})d\mu(\theta)V(e^{j\theta})^*. \tag{18}$$

The rows of $G_r(z)$, for $z = e^{j\theta}$, span

$$\mathcal{K}_r := \left(\mathcal{H}_2^{1\times m}\right)^\perp \ominus \left(\mathcal{H}_2^{1\times m}\right)^\perp V(z)^*.$$

The notation $^\perp$ denotes orthogonal complement in the "ambient" space—here $\mathcal{L}_2(\partial\mathbb{D})^{1\times m}$. It readily follows that a state-covariance of (2) satisfies the dual conditions given here.

Theorem 2: Let (4a-d) hold. A nonnegative-definite Hermitian matrix $\mathbf{R} \in \mathbb{C}^{n\times n}$ arises as the (stationary) state-covariance of (2) for a suitable stationary input process u_k if and only if the following equivalent conditions hold:

(5c) $\quad \mathrm{rank}\begin{bmatrix} \mathbf{R} - A^*\mathbf{R}A & C^* \\ C & 0 \end{bmatrix} = 2m$

or, equivalently

(5d) $\quad \mathbf{R} - A^*\mathbf{R}A = C^*L^* + LC$

for some $L \in \mathbb{C}^{n\times m}$

and C selected as in Section III so that $D + zC(I_n - zA)^{-1}B$ is inner. Conditions (5c-d) are equivalent to (5a-b).

It is noted that, $\mathrm{rank}(B) = m$ in condition (4b) implies that $\mathrm{rank}(C) = m$ as well. To see this, assume without loss of generality that (4e) holds. Then $B^*B = I_m - D^*D > 0$ which implies that $\|D\| < 1$. Using once more unitarity of U and the fact that $\|D\| < 1$, we obtain that $CC^* = I_m - DD^* > 0$ which implies that $\mathrm{rank}(C) = m$.

An insightful derivation of Theorem 2 can be obtained by considering (2) under time-reversal. More specifically, we compare the state-equations for dynamical systems with transfer functions $V(z) = D + Cz(I_n - zA)^{-1}B$ and $V(z)^* = D^* + B^*(zI_n - A^*)^{-1}C^*$ given as follows:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_k \\ y_k &= Cx_{k-1} + Du_k, k = \ldots, -1, 0, 1, \ldots \end{aligned} \tag{19}$$

and

$$\begin{aligned} x_{k-1} &= A^* x_k + C^* y_k \\ u_k &= B^* x_k + D^* y_k, k = \ldots, 1, 0, -1, \ldots . \end{aligned} \tag{20}$$

Both are interpreted as stable linear dynamical systems but with opposite time-arrows. Since $V(z)V(z)^* = I_m$, the input to one of the two corresponds to the output of the other, and (18) relates the spectral measure $d\mu$ of $\{u_k\}$ to the spectral measure $d\mu_r$ of $\{y_k\}$. The state-covariance for both systems is the same when the first is driven by $\{u_k\}$ and the second by $\{y_k\}$, respectively. Thus, if $\mathbf{R} = E\{x_k x_k^*\}$, Theorem 1 applied to (19) leads to (5a–b) while, applied to (20), leads to (5c-d). The spectral measures of the respective inputs $\{u_k\}$ and $\{y_k\}$ relate as in (18).

Proof: [Theorem 2]: Follows readily from the above arguments. More precisely, $\mathbf{R}$ is a state-covariance of (2) for a suitable stationary input process $\{u_k\}$ if and only if it is also a state-covariance of $x_{\ell+1} = A^* x_\ell + C^* y_\ell$ for a suitable stationary input process $\{y_\ell, \ell \in \mathbb{Z}\}$. Then applying Theorem 1 we draw the required conclusion. ■

An analogous dual interpolation problem ensues. To avoid repeat of the development in [19], [20], we may simply rewrite (17) as

$$\mathbf{R}' = \int_0^{2\pi} \left((G_r(e^{j\theta})^*)' \frac{(d\mu_r(\theta))'}{2\pi} G_r(e^{j\theta})' \right)$$

where now the left integration kernel is $(G_r(z)^*)' = (z^{-1}I_n - A')^{-1}C'$. Note that $\mathbf{R}' = \bar{\mathbf{R}} \neq \mathbf{R}$ in general, since $\mathbf{R}$ is Hermitian but may not be symmetric—where bar $(\bar{\ })$ denotes complex-conjugation. Trading a factor z between the left integration kernel and its para-hermitian conjugate on the right we obtain that

$$\mathbf{R}' = \int_0^{2\pi} (I_n - e^{j\theta}A')^{-1} C' \frac{(d\mu_r(\theta))'}{2\pi} C(I_n - e^{-j\theta}A)^{-1}$$

leading to the analytic interpolation problem of seeking an $\mathbb{F}$-function of the form

$$L'(I_n - zA')^{-1}C' + Q(z)V(z)'.$$

Transposing once more we may define

$$F_r(z) = C(I_n - zA)^{-1}L + V(z)Q(z)$$

and draw the following conclusions.

Theorem 3: Let $V(z) = D + Cz(I_n - zA)^{-1}B$ be an $m \times m$ inner function with (A, C) observable and (A, B) reachable. If $L \in \mathbb{C}^{n \times m}$ and $\mathbf{R}$ the solution to (5d), then there exists a solution H to (5b). Conversely, if $H \in \mathbb{C}^{m \times n}$ and $\mathbf{R}$ the solution to (5b), then there exists a solution L to (5d). With $\mathbf{R}, L, H$ related via (5b) and (5d), the following are equivalent:

(21a) $\mathbf{R} \geq 0$ (21)

(21b) $\exists F(z) \in \mathbb{F}:$ with $Q(z)$ analytic in $\mathbb{D}$
$F(z) = H(I - zA)^{-1}B + Q(z)V(z)$

(21c) $\exists F_r(z) \in \mathbb{F}:$ with $Q_r(z)$ analytic in $\mathbb{D}$
$F_r(z) = C(I - zA)^{-1}L + V(z)Q_r(z).$

Proof: Begin with $L \in \mathbb{C}^{n \times m}$ and $\mathbf{R}$ the solution to (5d). If $\mathbf{R} \geq 0$ then $\mathbf{R}$ is a state-covariance to (2) according to Theorem 2 and hence, there exists a solution H to (5b). To argue the case where $\mathbf{R}$ may not be nonnegative definite necessarily, consider without loss of generality condition (4e) valid and that U as in (10) is unitary. Then, $I_n - A^*A = C^*C$ and $I_n - AA^* = BB^*$. If $\mathbf{R}$ is the solution to (5d) for a given L, then $\mathbf{R}_\epsilon := \mathbf{R} + \epsilon I_n$ is the solution of the same equation when L is replaced by $L_\epsilon := L + (\epsilon/2)C^*$. We can always choose ϵ so that $\mathbf{R}_\epsilon > 0$ and then deduce that there exists a solution H_ϵ to

$$\mathbf{R}_\epsilon - A\mathbf{R}_\epsilon A^* = BH_\epsilon + H_\epsilon^* B^*. \tag{22}$$

Since $I_n = AA^* + BB^*, H := H_\epsilon - (\epsilon/2)B^*$ now satisfies (5b). The converse proceeds in the same way.

The equivalence of (21a) and (21b) follows as in [19]. If $\mathbf{R} \geq 0$, then (21b) follows from [19, Th. 2]. Conversely, if (21b) holds, then $\mathcal{W}$ [defined in (14)] satisfies (15) leading to $\mathbf{R}$ being its Hermitian part. Since $F(z) \in \mathbb{F}$, the Hermitian part of multiplication by $F(z)^*$ is nonnegative, and hence it remains so when restricted to the subspace $\mathcal{K}$.

The dual statement (21c) follows similarly. ■

Remark 1: If $V(z) = V_1(z)V_2(z)$ is a factorization of $V(z)$ into a product of inner factors, then it can similarly be shown that the conditions (21) of the theorem are equivalent to the solvability of a bitangential Carathéodory–Fejér interpolation problem of seeking an $F_o(z) \in \mathbb{F}$ where $F_o(z) = H_o(I - zA)^{-1}L_o + V_2(z)Q(z)V_1(z)$ for suitable H_o, L_o. The H_o, L_o can be computed from e.g., H, B by setting $F_o(z)$ as the analytic part of $V_2(z)F(z)V_2(z)^*$ and $F(z)$ as in (21b). An alternative treatment can be based on [2] or, in a more general setting, on [1]. □

V. Optimal Prediction and Postdiction Errors

A spectral distribution $d\mu \in \mathbb{M}$ induces a Gram matricial structure on the space of $p \times m$ matrix-valued functions on the circle (see [30, pp. 353, 361]) via

$$\langle a(z), b(z) \rangle_{d\mu} := \int_0^{2\pi} b(e^{j\theta}) \frac{d\mu(\theta)}{2\pi} a(e^{j\theta})^* \tag{23}$$

$$= \mathcal{E}\left\{ \left(\sum_\ell b_\ell u_{k-\ell} \right) \left(\sum_\ell a_\ell u_{k-\ell} \right)^* \right\} \tag{24}$$

where a_ℓ, b_ℓ are the Laurent coefficients of $a(z), b(z)$, respectively. The correspondence

$$\sum_\ell a_\ell z^\ell \mapsto \sum_\ell a_\ell u_{-\ell} \tag{25}$$

between functions on the unit circle (taking $z = e^{j\theta}$) and linear combinations of the random vectors $\{u_k\}$, leaves the respective Gram-matricial inner products in agreement and establishes a natural isomorphism between $\mathcal{L}_2(\partial\mathbb{D}; d\mu)$ and the space spanned by (the closure of) linear combination $\{u_k\}$ (see [30, Secs. 5 and 6], cf. [20]).

Any matrix-valued function $h(z) = \sum_{\ell=0}^{\infty} h_\ell z^\ell$ with entries in $\mathcal{H}_2$ and

$$h(0) = h_0 = I_m \tag{26}$$

corresponds via (25) to

$$h(z) \mapsto u_0 - \sum_{\ell=1}^{\infty} h_\ell u_{-\ell} =: u_0 - \hat{u}_{0|\text{ past}}$$

which is interpreted as a "one-step-ahead prediction error." Likewise, if $h(z) = \sum_{\ell=0}^{-\infty} h_\ell z^\ell \in z\mathcal{H}_2^\perp$ and $h_0 = I_m$

$$h(z) \mapsto u_0 - \sum_{\ell=1}^{\infty} h_\ell u_\ell =: u_0 - \hat{u}_{0|\text{ future}}$$

corresponds to "one-step-ahead postdiction error," i.e., using "future" observations only to determine the "present". Occasionally we may refer to these for emphasis as prediction forward, and backwards in time, respectively. Either way, the "estimator," which may not be optimal in any particular way, is the respective linear combination of values of u_k for $k \gtrless 0$

$$\hat{u}_{0|\text{ observation range}} := -\sum_{\ell \gtrless 0} h_\ell u_{-\ell}.$$

When the values extend in both directions it is a case of smoothing and is needed to interpret the $\mathbb{F}$-function in Remark 1—this will be developed in a forthcoming report.

We first discuss prediction in the forward direction. Throughout, we consider as data the covariance matrix $\mathbf{R}$ and the filter parameters. We assume that $d\mu \in \mathbb{M}_{\mathbf{R}}$ but otherwise unknown. Because $d\mu$ is not known outside $\mathcal{K}$, it can be shown that the min-max problem of identifying the forward prediction error with the least variance over all $d\mu \in \mathbb{M}_{\mathbf{R}}$ has a solution which lies in $\mathcal{K}$. To this end we seek an element in $\mathcal{K}^m$, i.e., an $m \times n$ matrix-valued function

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

with rows in $\mathcal{K}$, having least variance

$$\begin{aligned}\langle \Gamma G(z), \Gamma G(z)\rangle_{d\mu} &= \Gamma \langle G(z), G(z)\rangle_{d\mu} \Gamma^* \\ &= \Gamma \mathbf{R} \Gamma^*\end{aligned}$$

and subject to the constraint (26) which becomes

$$\Gamma B = I. \tag{27}$$

Existence and characterization of minimizing matrices Γ is discussed next.

Nonnegative definiteness of the difference $\Omega_1 - \Omega_2 \geq 0$ between two elements $\Omega_i \in \mathbb{H}_m (i = 1, 2)$ defines a partial order $\Omega_1 \geq \Omega_2$ in $\mathbb{H}_m$. An $\mathbb{H}_m$-valued function on a linear space is said to be $\mathbb{H}_m$-*convex* iff

$$f(\alpha\Gamma_1 + (1-\alpha)\Gamma_2) \leq \alpha f(\Gamma_1) + (1-\alpha) f(\Gamma_2) \quad \text{for } \alpha \in [0,1] \text{ and } \Gamma_1, \Gamma_2 \in \mathbb{C}^{m\times n}.$$

It is rather straightforward to check that if $\mathbf{R} \geq 0$, then

$$q_{\mathbf{R}} : \mathbb{C}^{m\times n} \to \mathbb{H}_m : \Gamma \mapsto \Omega = \Gamma \mathbf{R} \Gamma^* \tag{28}$$

is in fact $\mathbb{H}_m$-convex. This basic fact ensures existence of $\mathbb{H}_m$-minimizers satisfying (27) in the proposition given below.

Note that the statements ii) and iii) of the proposition are rephrased in alternative ways ii-a) and iii-a), in order to highlight an apparent symmetry. This is most clearly seen when expressed in terms of the two directed gaps $\vec{\delta}$ between the null space

$$\mathcal{N}(\mathbf{R}) := \{x \in \mathbb{C}^{n\times 1} : \mathbf{R}x = 0_n\}$$

of $\mathbf{R}$ and the range

$$\mathcal{R}(B) := \{x \in \mathbb{C}^{n\times 1} : x = Bv \text{ for } v \in \mathbb{C}^{m\times 1}\}$$

of B—the directed gap $\vec{\delta}$ is defined in the statement of the proposition and represents an angular distance between subspaces and is a standard tool in perturbation theory of linear operators (see [28]) and in robust control (e.g., see [23]).

Proposition 1: Let $B \in \mathbb{C}^{n\times m}$ having rank m, and let $\mathbf{R} \in \mathbb{H}_n$ with $\mathbf{R} \geq 0$. The following hold.

i) There exists an $\mathbb{H}_m$-minimizer of $q_{\mathbf{R}}$ satisfying (27).

ii) The minimizer is unique if and only if

$$\text{rank}([\mathbf{R} \quad B]) = n.$$

ii-a) The minimizer is unique if and only if

$$\vec{\delta}(\mathcal{N}(\mathbf{R}), \mathcal{R}(B)) := \left\| \mathbf{\Pi}_{\mathcal{R}(B)^\perp} |_{\mathcal{N}(\mathbf{R})} \right\| < 1.$$

iii) The $\mathbb{H}_m$-minimal value for $q_{\mathbf{R}}$ is O_m if and only if

$$B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} B \ \text{ is invertible.}$$

iii-a) The $\mathbb{H}_m$-minimal value for $q_{\mathbf{R}}$ is O_m if and only if

$$\vec{\delta}(\mathcal{R}(B), \mathcal{N}(\mathbf{R}))) := \left\| \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})^\perp} |_{\mathcal{R}(B)} \right\| < 1.$$

iv) If $\text{rank}(\mathbf{R}) = n$, then the $\mathbb{H}_m$-minimal value of $q_{\mathbf{R}}$ is

$$\Omega := (B^* \mathbf{R}^{-1} B)^{-1} > 0$$

and a minimizer [unique by ii)] is

$$\Gamma = (B^* \mathbf{R}^{-1} B)^{-1} B^* \mathbf{R}^{-1}.$$

v) If the $\mathbb{H}_m$-minimal value of $q_{\mathbf{R}}$ is $\Omega = O_m$, then a minimizer is given by

$$\Gamma = \left(B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} B\right)^{-1} B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}. \tag{29}$$

vi) In general, when $\mathbf{R}$ is singular, the $\mathbb{H}_m$-minimal value for $q_{\mathbf{R}}$ is

$$\Omega = (B_1^* \mathbf{R}^\sharp B_1)^\sharp \tag{30}$$

and a minimizer is given by

$$\begin{aligned}\Gamma = \left(B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} B\right)^\sharp B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} + (B_1^* \mathbf{R}^\sharp B_1)^\sharp B_1^* \mathbf{R}^\sharp \\ \times \left(I_n - B \left(B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} B\right)^\sharp B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}\right)\end{aligned} \tag{31}$$

where $\mathbf{R}^\sharp$ denotes the Moore–Penrose pseudoinverse

$$\begin{aligned} B_1 &:= B \mathbf{\Pi}_{\mathcal{N}_o}, \quad \text{and} \\ \mathcal{N}_o &:= \mathcal{N}\left(B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} B\right). \end{aligned}$$

Alternatively

$$\Omega = \lim_{\epsilon\to 0} \Omega_\epsilon \quad \text{and} \quad \Gamma = \lim_{\epsilon\to 0} \Gamma_\epsilon \tag{32}$$

where

$$\begin{aligned}\Omega_\epsilon &:= (B^*\mathbf{R}_\epsilon^{-1}B)^{-1}\\ \Gamma_\epsilon &:= \Omega_\epsilon B^*\mathbf{R}_\epsilon^{-1}\\ \mathbf{R}_\epsilon &:= \mathbf{R} + \epsilon\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}, \quad \text{and} \quad \epsilon > 0.\end{aligned} \tag{33}$$

Proof: Due to page limitations of this journal, the proof is omitted. However, it can be found in the full report which has been archived at http://arxiv.org/abs/math/0509225/ ■

Remark 2: It should be noted that $\mathbf{R}$ is not required to have the structure of a state-covariance of a reachable pair (A, B) (cf. Theorem 1) since the matrix A does not enter at all in the statement of Proposition 1. However, if this is the case (see Proposition 2) and $\mathbf{R}$ is a singular *state-covariance*, then Ω is singular as well—a converse to the first part of statement iv). Finally, $\mathbf{R}_\epsilon$ in (33) can also be taken as simply the perturbation $\mathbf{R} + \epsilon I_n$, though this complicates the algebra. □

For prediction backwards in time, the postdiction error

$$u_0 - \hat{u}_{0|\text{ future}} = u_0 - \sum_{\ell=1}^{\infty} h_\ell u_\ell \tag{34}$$

corresponds to an element

$$zL^*G_r(z) = L^*(I_n - z^{-1}A^*)^{-1}C^* \in z\mathcal{K}_r \quad \text{with } L \in \mathbb{C}^{n\times m}.$$

The constraint arising from the the identity in front of u_k in (34), translates into $L^*C^* = I_m$ while the variance of the postdiction error becomes $L^*\mathbf{R}L$. Proposition 1 applies verbatim and yields the following.

i') There exists an $\mathbb{H}_m$-minimal postdiction error.

ii') The minimizer is unique if and only if

$$\text{rank}\left(\begin{bmatrix}\mathbf{R}\\ C\end{bmatrix}\right) = n.$$

iii') The variance of optimal postdiction error is equal to O_m if and only if $C\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}C^*$ is invertible.

iv') If $\text{rank}(\mathbf{R}) = n$, then the variance of the optimal postdiction error is (strictly) positive definite and the unique minimizer is $\Gamma_r = \mathbf{R}^{-1}C^*(C\mathbf{R}^{-1}C^*)^{-1}$.

v') If the variance of the optimal postdiction error is equal to O_m, then a (non-unique) minimizer is $\Gamma_r = \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}C^*(C\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}C^*)^{-1}$.

Similarly, the analog of vi) holds as well.

Remark 3: It is interesting to point out that the square-roots of the variances of prediction and postdiction errors $(B^*\mathbf{R}^{-1}B)^{-1/2}$ and $(C\mathbf{R}^{-1}C^*)^{-1/2}$ appear as left and right radii, respectively, in a Schur parametrization of the elements of $\mathbb{M}_\mathbf{R}$ in [19] (cf. [20, Rem. 2]) and that, in view of the above, if one is zero so is the other. □

VI. When $\mathbb{M}_\mathbf{R}$ Contains a Single Element

We now focus on the case where $\mathbb{M}_\mathbf{R}$ consists of a single element, we analyze the nature of this unique power spectrum, and study ways to decompose $\mathbf{R}$ into a sum of two non-negative definite matrices, one of which has this property and another which may be interpreted as corresponding to noise. Conditions for $\mathbb{M}_\mathbf{R}$ to be a singleton are stated next.

Theorem 4: Let A, B satisfy (4) and $\mathbf{R} \geq 0$ for which (5) holds. Then, the set $\mathbb{M}_\mathbf{R}$ is a singleton if and only if the following equivalent conditions hold:

$$\begin{aligned}&\text{(35a)} \quad \vec{\delta}(\mathcal{R}(B), \mathcal{N}(\mathbf{R})) < 1\\ &\text{(35b)} \quad B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B \ \text{ is invertible.}\end{aligned} \tag{35}$$

If (C, D) are selected so that $V(z)$ in (11) is inner, the previous conditions are also equivalent to

$$\begin{aligned}&\text{(35c)} \quad \vec{\delta}(\mathcal{R}(C^*), \mathcal{N}(\mathbf{R})) < 1\\ &\text{(35d)} \quad C\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}C^* \ \text{ is invertible.}\end{aligned}$$

Proof: ($\Leftarrow$) We first assume that (35) holds (and hence, from Proposition 1, that (35a–d) are all valid). We recall the definition of C, D and $V(z)$ in (11) so that the latter is inner, as well as the definitions of $\mathcal{K} \subset \mathcal{H}_2^{1\times m}$ in (12), $G(z)$ in (7) whose rows form a basis for $\mathcal{K}$, and we now consider an arbitrary $d\mu \in \mathbb{M}_\mathbf{R}$. We show that $d\mu$ is unique and completely specified by the data $A, B, \mathbf{R}$.

As we have seen earlier, the quadratic form on $\mathcal{K}$ specified by $d\mu$ and expressed with respect to the rows of $G(z)$ as basis, is given by $\mathbf{R}$. Next we show that the quadratic forms specified by $d\mu$ on the nested sequence of subspaces $\mathcal{K} \subset \mathcal{K}_1 \subset \mathcal{K}_2 \subset \ldots$, with

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

for $\ell = 1, 2, \ldots$, are uniquely specified by $A, B, \mathbf{R}$. This is shown by induction. We first consider $\ell = 1$. If $V(z)$ is given by (11) then $zV(z)$ can be written as

$$V_1(z) := zV(z) = \mathcal{D}_1 + \mathcal{C}_1 z(I - z\mathcal{A}_1)^{-1}\mathcal{B}_1$$

with

$$\begin{aligned}\mathcal{A}_1 &= \begin{bmatrix}A & 0\\ C & 0\end{bmatrix}\\ \mathcal{B}_1 &= \begin{bmatrix}B\\ D\end{bmatrix}\\ \mathcal{C}_1 &= [0 \quad I]\\ \mathcal{D}_1 &= 0.\end{aligned} \tag{36}$$

The quadratic form induced by $d\mu$ with respect to the rows of $G_1(z) := (I - z\mathcal{A}_1)^{-1}\mathcal{B}_1$ is similarly given by

$$\mathbf{R}_1 := \begin{bmatrix}\mathbf{R} & \mathbf{R}_{1,12}\\ \mathbf{R}_{1,12}^* & \mathbf{R}_{1,22}\end{bmatrix}.$$

From Theorem 1

$$\mathbf{R}_1 - \mathcal{A}_1\mathbf{R}_1\mathcal{A}_1^* = \mathcal{B}_1\mathcal{H}_1 + \mathcal{H}_1^*\mathcal{B}^* \tag{37}$$

where

$$\mathcal{H}_1 = [H \quad H_1]. \tag{38}$$

Let Γ be as in (29) (for which $\Gamma\mathbf{R} = O_{m\times n}$ and $\Gamma B = I_m$). Since, $\mathbf{R}_1 \geq 0$, it follows that $\Gamma\mathbf{R}_{1,12} = O_{m\times m}$, otherwise it would be possible to render the quadratic form $\alpha\Gamma\mathbf{R}_{1,12} + \bar{\alpha}\mathbf{R}_{1,12}^*\Gamma^* + |\alpha|^2\mathbf{R}_{1,22}$ indefinite with a suitable choice of $\alpha \in \mathbb{C}$ which would contradict $\mathbf{R}_1 \geq 0$. From (37) on the other hand, we have that $\mathbf{R}_{1,12} - A\mathbf{R}C^* = BH_1 + H^*D^*$. Multiplying on the left by Γ we conclude that $H_1 = -\Gamma A\mathbf{R}C^* - \Gamma H^*D^*$ is uniquely specified from the original data $A, B, \mathbf{R}$. In view of (37) the same is true for the "one-step" extension $\mathbf{R}_1$ of $\mathbf{R}$. In order to proceed with the induction, we only need to show that the condition (35a) continues to be valid for the new data, i.e., that $\mathcal{B}_1^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R}_1)}\mathcal{B}_1$ is also invertible. To prove this last claim we argue as follows. Since $\mathbf{R}$ is Hermitian, $\mathbb{C}^n = \mathcal{N}(\mathbf{R}) \oplus \mathcal{R}(\mathbf{R})$ is an orthogonal decomposition. Then, the null space of $\mathbf{R}_1$ is the orthogonal direct sum of

$$\mathcal{N}_1 := \left\{ \begin{pmatrix} \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}x \\ 0 \end{pmatrix} : x \in \mathbb{C}^n \right\}$$

and

$$\mathcal{N}_2 := \left\{ \xi = \begin{pmatrix} \mathbf{\Pi}_{\mathcal{R}(\mathbf{R})}x \\ y \end{pmatrix} : x \in \mathbb{C}^n, y \in \mathbb{C}^m, \mathbf{R}_1\xi = 0 \right\}.$$

Then, $\mathbf{\Pi}_{\mathcal{N}(\mathbf{R}_1)} = \mathbf{\Pi}_{\mathcal{N}_1} + \mathbf{\Pi}_{\mathcal{N}_2}$ where $\mathbf{\Pi}_{\mathcal{N}_1} = \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} \oplus O_{m\times m}$. So, finally, $\mathcal{B}_1^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R}_1)}\mathcal{B}_1 = B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B + \mathcal{B}_1^*\mathbf{\Pi}_{\mathcal{N}_2}\mathcal{B}_1 > 0$, because $B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B$ is already positive definite. This completes the proof of our claim that $\mathcal{B}_1^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R}_1)}\mathcal{B}_1$ is invertible. It also completes the general step of the inductive argument, i.e., it proves that if $\mathbf{R}_\ell$ specifies the quadratic form induced by $d\mu$ on $\mathcal{K}_\ell$ with respect to the rows of $G_\ell(z)$ as a basis, then $\mathbf{R}_\ell$ is uniquely specified by $A, B, \mathbf{R}$ for $\ell = 1, 2, \ldots$.

In order to prove that $d\mu$ is uniquely defined as well, we compute an infinite sequence of relevant Fourier coefficients and show that they are also specified uniquely by $A, B, \mathbf{R}$. To this end, we first note that

$$G_1(z) = \begin{bmatrix} G(z) \\ V(z) \end{bmatrix}.$$

Similarly

$$G_\ell(z) = \begin{bmatrix} G(z) \\ V(z) \\ zV(z) \\ \vdots \\ z^{\ell-1}V(z) \end{bmatrix}.$$

If we partition $\mathbf{R}_\ell = \int_{-\pi}^{\pi} G_\ell(e^{j\theta})d\mu(\theta)G_\ell(e^{j\theta})^*$ into blocks, accordingly, we observe that after omitting the (1, 2) entry the second column consists of the Fourier coefficients

$$\int_{-\pi}^{\pi} e^{jk\theta}V(e^{j\theta})d\mu(\theta)V(e^{j\theta})^* \quad \text{for } k = 0, 1, \ldots, \ell - 1$$

of $d\mu_r(\theta) = V(e^{j\theta})d\mu(\theta)V(e^{j\theta})^*$ [which was encountered earlier in (18)]. Since all the Fourier coefficients are uniquely specified by $A, B, \mathbf{R}$, so is $d\mu_r(\theta)$. This is due to the fact that the matricial trigonometric problem is determinate very much like in the scalar case [8, Th. 1]. Of course, $d\mu(\theta)$ is uniquely specified as well and therefore $\mathbb{M}_\mathbf{R}$ is a singleton.

($\Rightarrow$) If $\mathbf{R}_1$ is an extension of $\mathbf{R}$ as before, $\mathbb{M}_{\mathbf{R}_1} \subseteq \mathbb{R}_\mathbf{R}$. Furthermore, if $\mathbf{R}_1^\alpha \neq \mathbf{R}_1^\beta$ are two different extensions, $\mathbb{M}_{\mathbf{R}_1^\alpha} \cap \mathbb{M}_{\mathbf{R}_1^\beta} = \emptyset$. Thus, to prove that $\mathbb{M}_\mathbf{R}$ is not a singleton, it suffices to prove that there are choices in specifying H_1 while maintaining $\mathbf{R}_1 \geq 0$. It turns out that the condition $\mathbf{R}_1 \geq 0$ restricts H_1 to a matrix ball with nontrivial left and right radii. The computations have already been carried out in [19], [20] for the case where $\mathbf{R}$ is invertible. Here, we first summarize the relevant facts and conclude the proof using by a limit argument.

Without loss of generality we assume that A, B are normalized so that (4e) holds. We define $\mathbf{R}_\epsilon := \mathbf{R} + \epsilon I_n, H_\epsilon := H + (\epsilon/2)B^*$, and we define $\mathbf{R}_{1,\epsilon}$ via

$$\mathbf{R}_{1,\epsilon} - \mathcal{A}_1\mathbf{R}_{1,\epsilon}\mathcal{A}_1^* = \mathcal{B}_1\mathcal{H}_{1,\epsilon} + \mathcal{H}_{1,\epsilon}^*\mathcal{B}^* \tag{39}$$

taking

$$\mathcal{H}_{1,\epsilon} = [H_\epsilon \quad H_1]. \tag{40}$$

We note that the (1,1) entry of $\mathbf{R}_{1,\epsilon}$ is simply $\mathbf{R}_\epsilon$. With U as in (10), we compute

$$U^*\mathbf{R}_{1,\epsilon}U = \begin{bmatrix} \mathbf{R}_\epsilon & C^*H_1^* + A^*H_\epsilon^* \\ H_1C + H_\epsilon A & H_\epsilon B + B^*H_\epsilon^* + H_1D + D^*H_1^* \end{bmatrix}.$$

Since $\mathbf{R}_\epsilon > 0$, the non-negativity of $\mathbf{R}_{1,\epsilon}$ is equivalent to the non-negativity of the Schur complement of the 2×2-block matrix on the right-hand side of the previous equation, pivoted about its (1,1) entry. This leads to a quadratic expression for H_1 which needs to be non-negative. As shown in [19], (47)–(51) and in [20, Rem. 2], after standard (but rather complicated) algebraic manipulations we conclude that $\mathbf{R}_{1,\epsilon} \geq 0$ if and only if

$$H_1 \in \{c_\epsilon + r_{\text{left},\epsilon}sr_{\text{right},\epsilon} : s \in \mathbb{C}^{m\times m}, \|s\| \leq 1\} \tag{41}$$

where the center of the ball is

$$c_\epsilon = D^*(C\mathbf{R}_\epsilon{}^{-1}C^*)^{-1} - HA\mathbf{R}_\epsilon{}^{-1}C^*(C\mathbf{R}_\epsilon{}^{-1}C^*)^{-1}$$

and two radii are

$$r_{\text{left},\epsilon} = (B^*\mathbf{R}_\epsilon^{-1}B)^{-1/2}, \quad \text{and}$$
$$r_{\text{right},\epsilon} = (C\mathbf{R}_\epsilon^{-1}C^*)^{-1/2}.$$

Following part (iv) of Proposition 1, the columns of $B_1 = B\mathbf{\Pi}_{\mathcal{N}_o}$ where $\mathcal{N}_o = \mathcal{N}(B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B) \neq \emptyset$, from a basis for $\mathcal{R}(\mathbf{R}) \cap \mathcal{R}(B) \neq \emptyset$. Similarly for $C_1 = \mathbf{\Pi}_{\mathcal{N}_1}C$ with $\mathcal{N}_1 = \mathcal{N}(C\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}C^*) \neq \emptyset$ is nonzero with the columns of C_1^* in the range of $\mathbf{R}$. Then, as $\epsilon \searrow 0$, the limit of the two radii turn out to be

$$r_{\text{left}} = ((B_1^*\mathbf{R}^\sharp B_1)^\sharp)^{1/2}, \quad \text{and}$$
$$r_{\text{right}} = ((C_1\mathbf{R}^\sharp C_1^*)^\sharp)^{1/2}$$

while a corresponding expression can be obtained for the limit of c_ϵ as well involving $C, C_1, \mathbf{R}$ and $\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}$. The fact that neither r_{left} nor r_{right} is zero implies that we can select at least two families of values for H_1 indexed by ϵ and belonging to the matrix ball in (41) for each ϵ, having different limits as $\epsilon \searrow 0$. Each choice renders $\mathbf{R}_{1,\epsilon} > 0$, and hence in the limit, $\mathbf{R}_1 \geq 0$ for the two different limit values for H_1. This completes the proof. ∎

The unique element in $\mathbb{M}_\mathbf{R}$ under the conditions of the theorem can be obtained, in principle, after extending $\mathbf{R}$ recursively for $\ell = 1, 2, \ldots$ using (36)–(38). This specifies a non-negative operator on a dense subset of $\mathcal{H}_2^{1\times m}$ which, in turn, specifies a corresponding positive real function $F(z)$ and the measure can be obtained from the boundary limits of the real part of $F(z)$ as a weak limit. However, an explicit expression for $F(z)$ will also be given later on. Before we do this, we explain some of the properties of this unique measure.

The following result states that $d\mu$ is a singular measure with at most $n-m$ points of increase, i.e., at most $n-m$ spectral lines whose directionality is encapsulated in suitably chosen unitary factors. The spectral lines are in fact at the zeros of certain matrix-valued functions, namely

$$\Phi(z) := \Gamma G(z) = \Gamma(I_n - zA)^{-1}B \tag{42}$$

and Γ as in Proposition 1, which correspond to the optimal prediction error and represent the analog of the Szegö–Geronimus orthogonal polynomials of the first kind, cf. [20].

Theorem 5: Under the assumptions and conditions of Theorem 4, the unique element in $\mathbb{M}_{\mathbf{R}}$ is of the form

$$d\mu(\theta) = \sum_{\ell=1}^{q} V_\ell \rho_\ell V_\ell^* d\mathbb{U}(\theta - \theta_\ell)$$

where $\sum_1^q \operatorname{rank}(V_\ell) \leq n - m, \theta_\ell \in [0, 2\pi)$ for $\ell = 1, \ldots, q$ differ from one another, $\mathbb{U}(\theta - \theta_\ell)$ denotes a unit step at θ_ℓ, and $\rho_\ell > 0$. The values $e^{j\theta_\ell}$ for $\ell = 1, \ldots, q$ are the non-zero eigenvalues of the matrix $(I_n - B\Gamma)A$ with Γ as in (29). The matrices V_ℓ are chosen so that

$$\mathcal{R}(V_\ell) = \mathcal{N}(B^* \mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}(I_n - e^{j\theta_\ell}A)^{-1}B)$$

and can be normalized to satisfy $V_\ell V_\ell^* = I$ as well as to make ρ_ℓ diagonal.

Proof: Under the stated conditions, $\mathbb{M}_{\mathbf{R}}$ is a singleton by the previous theorem and its unique element $d\mu$ satisfies

$$\int_0^{2\pi} (\Phi(e^{j\theta}) d\mu(\theta) \Phi(e^{j\theta})^*) = \Gamma \mathbf{R} \Gamma^* = O_{m\times m} \tag{43}$$

with Γ as in (29). It readily follows that $d\mu$ can have points of increase only at the finitely many points $\theta_\ell, \ell = 1, \ldots, q$, where $\Phi(e^{j\theta})$ is singular. The "zeros" of $\Phi(z)$ coincide with the "poles" of its inverse

$$\Phi(z)^{-1} = I_n - \Gamma A(I_n - zA_o)^{-1}B \tag{44}$$

where

$$A_o = (I_n - B\Gamma)A. \tag{45}$$

Since A_o has already m eigenvalues at the origin, the number of eigenvalues that it may have on the circle is at most $n - m$. Thus

$$d\mu(\theta) = \sum_{\ell=1}^{q} M_\ell d\mathbb{U}(\theta - \theta_\ell)$$

where $M_\ell \in \mathbb{H}_m, M_\ell \geq 0$, and $\Phi(e^{j\theta_\ell})M_\ell = O_{m\times m}$. Expressing $M_\ell = V_\ell \rho_\ell V_\ell^*$ with ρ_ℓ, V_ℓ as claimed is standard. This completes the proof. ■

Thus, $\mathbb{M}_{\mathbf{R}}$ being a singleton implies just as in the classical scalar case (e.g., [32] and [26]) the underlying stochastic process is deterministic with finitely many complex exponential components. Subspace identification techniques represent different ways to identify "dominant ones" and obtain the "residue" ρ_ℓ that corresponds to each of those modes in the scalar (see [26], [32], [17], [18]). In order to do something analogous for multivariable stochastic processes, we need an explicit expression for the corresponding positive real function [corresponding via (6) and (8)]. This is obtained in the next section.

Remark 4: A dual version of the representation in Theorem 5 gives that θ_ℓ correspond to "zeros" on the circle of the optimal postdictor error $L^*(zI_n - A^*)^{-1}C^*$. Similarly, the range $\mathcal{R}(V_\ell)$, for $\ell = 1, 2, \ldots, q$, is contained in the correspond null space of the previous postdiction error when evaluated at the corresponding zeros. □

Remark 5: The "star" of the optimal postiction error can also be interpreted as a "right matricial orthogonal polynomial of the first kind"

$$\Phi_r(z) = C(I_n - zA)^{-1}L. \tag{46}$$

These matricial functions, i.e., $\Phi(z)$ and $\Phi_r(z)$, together with their counterparts of the "second kind" $\Psi(z)$ and $\Psi_r(z)$ that will be introduced in the next section, satisfy a number of interesting properties similar to those of the classical orthogonal polynomials [24] (cf. [8] and [9]). We plan to develop this subject in a separate future publication. □

VII. The "Central" Positive Real Function

When $\mathbf{R} > 0$, there is a unique element in $\mathbb{M}_{\mathbf{R}}$ which maximizes the (concave) entropy functional

$$\mathbb{I}(\mu) := \int_0^{2\pi} \log\det(\dot{\mu}(\theta))\, d\theta. \tag{47}$$

This element of $\mathbb{M}_{\mathbf{R}}$ can be identified in a variety of ways (e.g., see [20], [21], also ([12], Chapter 11) and the references therein) as

$$d\mu_{\mathrm{ME}} = \Phi(e^{j\theta})^{-1}\Omega(\Phi(e^{j\theta})^{-1})^* d\theta \tag{48}$$

and generalizes a classical result in [24, eq. (1.20)]—the subscript ME suggesting "maximum entropy." However, in general, when $\mathbf{R} \geq 0$ and singular, (48) is no longer valid. Singular parts may be present in every element of $\mathbb{M}_{\mathbf{R}}$, reflecting purely deterministic components in the underlying time series. Yet the expression in (48) cannot reflect such deterministic components. Thus, when $\mathbf{R}$ is singular, and in order to identify power spectra, we need to make use of $\mathbb{F}$-functions corresponding to elements in $\mathbb{M}_{\mathbf{R}}$ via (8). The singular part of the power spectral measure can then be recovered either through (9) or by isolating the contribution of poles of the $\mathbb{F}$-function which lie on the boundary of the circle, as it will be explained in Section VIII.

Therefore, the purpose of this section is to present an expression for $F_{\mathrm{ME}}(z) = \mathcal{H}[\mu_{\mathrm{ME}}]$ which generalizes the scalar result in, e.g., [24, eq. (1.14)]. This expression remains valid when $\mathbf{R}$ is singular in that the corresponding measure $d\mu_{\rm ME} \in \mathbb{M}_{\mathbf{R}}$. However, $d\mu_{\rm ME}$ neither satisfies (48) nor represents a maximizer of (47) (since the value of the functional in (47) may equal $-\infty$). Our approach, which is detailed in the proof of Theorem 6, has been to seek a fractional representation for $F_{\rm ME}(z)$ which directly generalizes the scalar case in ([24], (1.16)). Indeed, this turns out to be possible leading to the following.

With $\mathbf{R}, A, B, H$ satisfying (5b) in Theorem 1, define

$$F_{\rm ME}(z) := \Phi(z)^{-1}\Psi(z) \tag{49}$$

where $\Phi(z) = \Gamma(I_n - zA)^{-1}B$ as before

$$\Psi(z) := -\Gamma z(I_n - zA)^{-1}AH^* + D_\Psi \tag{50}$$

and

$$D_\Psi := -\Gamma(H^*B^* - \mathbf{R})B(B^*B)^{-1}. \tag{51}$$

Here, $\Psi(z)$ is the analog of the Szegö orthogonal polynomial of the *second kind* (see [24, eq. (1.13)]) and was constructed mimicking the scalar case, with the same dynamics as $\Phi(z)$ so that $F_{\rm ME}$ is positive real and consistent with $\mathbf{R}$. Both, consistency, which led to the specific expressions for D_Ψ and $B_\Psi := AH^*$, as well as membership in $\mathbb{F}$ are verified in the proof of Theorem 6.

Before we proceed with Theorem 6, we present an alternative expression for $F_{\rm ME}(z)$ which can be obtained via standard multivariable "pole-zero" cancellation in the fractional representation (49). Briefly, because $\Phi(z) = I_m + z\Gamma(I_n - zA)^{-1}B$ and $\Psi(z) = D_\Psi - z\Gamma(I_n - zA)^{-1}B_\Psi$ share the same "state-matrix" A and the same "read-out" map Γ, a state–space representation for the cascade connection of the two systems with transfer functions $\Phi(z)^{-1}$ and $\Psi(z)$ has unobservable dynamics. Multivariable "pole-zero" cancellation then amounts to obtaining a canonical realization for this cascade connection. Since the dynamics associated with the matrix A turn out to be unobservable, the canonical realization is of dimension n and can be simplified into

$$F_{\rm ME}(z) = D_\Psi + zC_o(I_n - A_o)^{-1}B_o \tag{52}$$

where

$$\begin{aligned} C_o &:= -\Gamma A \\ A_o &:= (I_n - B\Gamma)A \\ B_o &:= BD_\Psi + H^*. \end{aligned} \tag{53}$$

We now turn to establishing our claim that $F_{\rm ME}$ is positive real and consistent with $\mathbf{R}$.

Theorem 6: Let $\mathbf{R}, A, B, H$ satisfy (5b) of Theorem 1, Γ given as in (31), and $F_{\rm ME}(z)$ given as in (49)–(51). Then

i) $F_{\rm ME}(z)$ satisfies (16);

ii) $F_{\rm ME}(z) \in \mathbb{F}$.

Proof: Condition (16) is equivalent to

$$\Psi(z)V(z)^* - \Phi(z)HG(z)V(z)^* = \Phi(z)Q(z).$$

To show that this relationship holds for some $Q(z)$ analytic in $\mathbb{D}$, it suffices to show that all negative Fourier coefficients of

$$\Psi(z)V(z)^* - \Phi(z)HG(z)V(z)^* \tag{54}$$

vanish. By collecting positive and negative powers of z we can express

$$\begin{aligned}\Psi(z)V(z)^* = (D_\Psi B^* - \Gamma AWA^*)(zI_n - A^*)^{-1}C^* \\ + D_\Psi D^* - \Gamma A(I_n - zA)^{-1}(zH^*D^* + WC^*)\end{aligned}$$

and similarly that

$$\begin{aligned}\Phi(z)HG(z)V(z)^* = \Gamma W^*(zI_n - A^*)^{-1}C^* \\ + \Gamma(I_n - zA)^{-1}WC^*\end{aligned}$$

where W is given by (15). Thus, negative powers of z in (54) sum up into

$$(D_\Psi B^* - \Gamma(AWA^* + W^*))(zI_n - A^*)^{-1}C^*.$$

Thus, to prove our claim (and because (A^*, C^*) is reachable), we need to show that $D_\Psi B^* - \Gamma(AWA^* + W^*)$ vanishes. Substituting the value for D_Ψ from (51) in the above expression we get

$$\begin{aligned} &-\Gamma(H^*B^* - \mathbf{R}B(B^*B)^{-1}B^* + AWA^* + W^* \\ &\quad = -\Gamma(W - \mathbf{R}B(B^*B)^{-1}B^* + W^*) \\ &\quad = -\Gamma\mathbf{R}(I_n - B(B^*B)^{-1}B^*). \end{aligned}$$

Recall that $\Gamma = \lim_{\epsilon\to 0}\Gamma_\epsilon$, from the proof of Proposition 1, while Γ_ϵ satisfies

$$\Gamma_\epsilon\left(\mathbf{R} + \epsilon\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}\right) = \Omega_\epsilon^{-1}B^*.$$

Thus

$$\Gamma_\epsilon\left(\mathbf{R} + \epsilon\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}\right)(I_n - B(B^*B)^{-1}B^*) = 0$$

identically for all ϵ, and hence, taking the limit as $\epsilon \to 0$ we get the desired conclusion. This completes the proof of claim i).

We first argue that $F_{\rm ME}(z)$ is analytic in $\mathbb{D}$. Of course, $\Psi(z)$ is already analytic in $\mathbb{D}$ by our standing assumption on the location of the eigenvalues of A. (Its poles cancel with the corresponding zeros of $\Phi(z)^{-1}$ anyway.) We only need to consider $\Phi(z)^{-1}$. If $\mathbf{R}$ is invertible, then $\Phi(z)^{-1}$ has no poles in $\mathbb{D}$ by [20, Prop. 1]. If $\mathbf{R}$ is singular, then, once again, we consider

$$\mathbf{R}_\epsilon = \mathbf{R} + \epsilon\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}, \text{ with } \epsilon > 0.$$

With $\Omega_\epsilon = (B^*\mathbf{R}_\epsilon^{-1}B)^{-1}$ and $\Gamma_\epsilon = \Omega_\epsilon B^*\mathbf{R}_\epsilon^{-1}$ as before we define $\Phi_\epsilon(z) := \Gamma_\epsilon G(z)$ and apply [20, Prop. 1] to deduce that $\Phi_\epsilon(z)^{-1}$ is analytic in the closed unit disc, for all $\epsilon > 0$. By continuity, $\Phi(z)$ has no poles in the open unit disc. Similarly, the Hermitian part of $F_{\rm ME}(z)$ in $\mathbb{D}$ is the limit of the Hermitian part of

$$F_{{\rm ME},\epsilon}(z) := \Phi_\epsilon(z)^{-1}\Psi_\epsilon(z)$$

where $\Psi_\epsilon(z)$ is given by (50) with $\Gamma, \mathbf{R}$ replaced by $\Gamma_\epsilon, \mathbf{R}_\epsilon$, respectively. A matricial version of a classical identity between orthogonal polynomials (of first and second kind [24, eq. (1.17)]) holds here as well

$$\Psi(z)\Phi(z)^* + \Phi(z)\Psi(z)^* = \Omega. \tag{55}$$

To verify this, after standard algebraic rearrangement, the left-hand side becomes

$$\Lambda_0 + z\Gamma(I_n - zA)^{-1}B_+ + z^{-1}B_+^*(I_n - z^{-1}A^*)^{-1}\Gamma^*$$

where

$$\begin{aligned} B_+ &= A(BD_\Psi^* - \mathbf{R}\Gamma^* + BH\Gamma^*), \quad \text{and} \\ \Lambda_0 &= D_\Psi + D_\Psi^* - \Gamma A\mathbf{R}A^*\Gamma^*. \end{aligned}$$

If $\mathbf{R}$ is invertible it is straightforward to show that

$$\mathrm{BD}_\Psi^* - R\Gamma^* + BH\Gamma^* = O_{n,m}$$

while

$$\Lambda_0 = (B^*\mathbf{R}^{-1}B)^{-1} = \Omega.$$

If $\mathbf{R}$ is singular then, as usual, we replace $\Gamma, \mathbf{R}$ by their ϵ-perturbations and claim the same identities for the relevant limits. This shows that $F_{\mathrm{ME},\epsilon}(z) \in \mathbb{F}$ for all $\epsilon > 0$. Hence, so is $F_{\mathrm{ME}}(z)$ since it is analytic in $\mathbb{D}$ and its Hermitian part is nonnegative being the limit of the Hermitian part of $F_{\mathrm{ME},\epsilon}(z)$ as $\epsilon \to 0$. ■

Remark 6: The expression (49) for $F_{\mathrm{ME}}(z)$ can in principle be obtained in a variety of ways, e.g., using the linear fractional parametrizations of solutions to related analytic interpolation problems in, e.g., [20, Th. 2] and [2, Ch. 22], or developed in analogy with [12, Ch. 11], [13]. However, the singularity of $\mathbf{R}$ renders the necessary algebra quite challenging.

Remark 7: The relationship (55) (cf. [24, eq. (1.17)]) between matricial functions of the "first" and "second-kind" generalizes to a two-sided version. Indeed, if we introduce analogous quantities for a right fraction

$$F_{\mathrm{ME}}(z) = \Psi_r(z)\Phi_r(z)^{-1}$$

by taking $\Phi_r(z)$ as in (46) and

$$\begin{aligned} \Psi_r(z) &:= -L^*zA(I_n - zA)^{-1}\Gamma_r + D_{\Psi_r} \\ D_{\Psi_r} &== -(CC^*)^{-1}C(C^*L^* - \mathbf{R})\Gamma_r \end{aligned}$$

then these satisfy

$$\begin{aligned} &\begin{bmatrix} \Psi_\ell(z) & \Phi_\ell(z) \\ \Phi_r(z)^* & \Psi_r(z)^* \end{bmatrix} \begin{bmatrix} \Phi_r(z) & \Phi_\ell(z)^* \\ -\Psi_r(z) & \Psi_\ell(z)^* \end{bmatrix} \\ &\quad = \begin{bmatrix} O_m & \Omega_\ell \\ \Omega_r & O_m \end{bmatrix} \\ &\quad := \begin{bmatrix} O_m & (C\mathbf{R}^\sharp C^*)^\sharp \\ (B^*\mathbf{R}^\sharp B)^\sharp & O_m \end{bmatrix}. \end{aligned} \tag{56}$$

In the above we subscribe ℓ, setting $\Phi_\ell(z) = \Phi(z)$ and $\Psi_\ell(z) = \Psi(z)$, to highlight "left functions" since $\Phi(z), \Psi(z)$ are the entries of the left fraction $F_{\mathrm{ME}}(z) = \Phi(z)^{-1}\Psi(z)$ of $F_{\mathrm{ME}}(z)$. □

VIII. Multivariable "Residues"; and Singular Parts

We begin with

$$\begin{aligned} F(z) := F_{\mathrm{ME}}(z) &= \Phi(z)^{-1}\Psi(z) \\ &= D_\Psi + C_oz(I_n - zA_o)^{-1}B_o \end{aligned} \tag{57}$$

as given in (52), suppressing the subscript "ME" for convenience. When $\mathbf{R} > 0$, then $\Phi(z)$ remains invertible in the closed unit disc and (55) readily implies that

$$\mathrm{Herm}\{F(e^{j\theta})\} = \Phi(e^{j\theta})^{-1}\Omega\left(\Phi(e^{j\theta})^{-1}\right)^* \tag{58}$$

cf. (48). However, when $\mathbf{R}$ is singular, the variance of the minimal prediction error Ω is also singular (see Proposition 2) and (48) may no longer be valid. The boundary limit of the Hermitian part defines a measure which may no longer be absolutely continuous. However, because $F(z)$ is rational the singular part consists of finitely many disconinuities in $\mu(\theta)$. In order to separate the singular part from the absolutely continuous, we need to isolate the boundary poles of $F(z)$. Accordingly, $F(z)$ decomposes into a sum of "lossless" and "lossy" components—the lossless part being responsible for the singular part of the measure.

In the case where $F(z) \in \mathbb{F}$ is scalar-valued, the multiplicity of any pole

$$\xi \in \partial\mathbb{D} := \{z : \|z\| = 1\}$$

cannot exceed one and $F(z)$ decomposes into

$$\rho\left(\frac{1 + z/\xi}{1 - z/\xi}\right) + F_{\mathrm{remaining}}(z) \text{ with } \rho > 0$$

where the first term is "lossless" and the second, $F_{\mathrm{remaining}}(z) \in \mathbb{F}$, has no singularity at ξ. Conformably

$$d\mu(\theta) = \rho d\mathbb{U}(\theta - \angle\xi) + d\mu_{\mathrm{remaining}}(\theta)$$

where $\angle\xi$ denotes the angle of ξ (i.e., $\xi = e^{j\angle\xi}$) and $d\mu_{\mathrm{remaining}}(\theta)$ is continuous at $\angle\xi$. Thus, in general

$$F(z) = \sum_{\ell=1}^{q} \rho_i \left(\frac{1 + z/\xi_i}{1 - z/\xi_i}\right) + F_{\mathrm{lossy}}(z)$$

and the corresponding measure

$$d\mu(\theta) = \sum_{i=1}^{q} \rho_i\, d\mathbb{U}(\theta - \theta_i) + \dot{\mu}(\theta)\, d\theta.$$

Analogous facts hold true in the multivariable case with some exceptions. Singularities in $\mathbf{R}$ may not necessarily be associated with discontinuities in the measure and, while $F(z)$ can have poles with higher multiplicity on the boundary of $\mathbb{D}$, these may not have geometric multiplicity exceeding one. When $F(z)$ has poles on the boundary, these are associated with discontinuities and our interest is to show how to decompose $F(z)$ into a lossless and a lossy part, in general, and thus isolate the singular part of the measure. We first discuss the significance of $\mathbf{R}$ being singular. With A_o, B_o, C_o as in (53) and Γ, Ω as in Proposition 1 it holds that

$$\mathbf{R} = B\Omega B^* + A_o\mathbf{R}A_o^*. \tag{59}$$

This can be verified directly (by careful algebra). It can also be shown via a limiting argument, replacing $\mathbf{R}, \Gamma, A_o$ with $\mathbf{R}_\epsilon, \Gamma_\epsilon, (I_n - B\Gamma_\epsilon)A$ (as in the proof of Theorem 6) and invoking [20, eq. (23)] to show that a similar identity holds for the perturbed quantities for all $\epsilon > 0$, hence for their limits as well. A direct consequence of (59) is the following.

Proposition 2: Let A, B satisfy (4a-d), $\mathbf{R} \geq 0, A, B, \mathbf{R}$ satisfy (5), and Ω the $\mathbb{H}_m$-minimal value of $q_{\mathbf{R}}$ subject to (27). If $\Omega > 0$, then $\mathbf{R} > 0$.

Proof: The pair $(A_o, B\Omega^{1/2})$ is a reachable pair since it is obtained from (A, B) after a state-feedback transformation and an invertible input tranformation. Then, $\mathbf{R}$ must be the reachability Grammian from (59) which cannot be singular. ■

Example 1: Elementary scalar examples suffice to demonstrate how singularities of $\mathbf{R}$ can give rise to poles of $F(z)$ on $\partial\mathbb{D}$. To see that this may not always be the case consider A, B as in (3) with $n = 4$ and $m = 2$, and let $\mathbf{R}$ which is now block-Toeplitz as in (1) have entries

$$R_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad R_1 = \frac{1}{2}R_0.$$

Then

$$\Gamma = \begin{bmatrix} I_2 & -\frac{1}{4}R_0 \end{bmatrix} \quad \text{and} \quad \Omega = \frac{3}{4}R_0.$$

Both $\mathbf{R}$ and Ω are singular while the eigenvalues of A_o are $\{0, 0, 0, (1/2)\}$. □

Next, we present some general facts about lossless rational matrices in $\mathbb{F}$. If $d\mu(\theta) = \sum_{\ell=1}^{q} V_\ell \rho_\ell V_\ell^* d\mathbb{U}(\theta - \theta_\ell)$ then

$$\mathcal{H}[d\mu(\theta)] = D_s + C_s z(I_{n_s} - zA_s)^{-1}B_s =: F_s(z)$$

where

$$\begin{aligned} D_s &= \sum_{\ell=1}^{q} V_\ell \rho_\ell V_\ell^* \\ C_s &= 2[e^{j\theta_1}V_1 \quad \ldots \quad e^{j\theta_q}V_q] \\ B_s &= \begin{bmatrix} \rho_1 V_1^* \\ \vdots \\ \rho_q V_q^* \end{bmatrix} \end{aligned}$$

and A_s block diagonal with blocks of the form $e^{j\theta_\ell}I_{n_\ell}$ of size equal to the size of ρ_ℓ. Then $F_s(s) \in \mathbb{F}$ but it is also *lossless*, which amounts to $\mathrm{Herm}\{F_s(re^{j\theta})\} = 0$ a.e. on $\partial\mathbb{D}$. It is a consequence of the Herglotz representation that, modulo a state transformation and an additive skew-Hermitian summand in D, any rational lossless function is necessarily of this form. An alternative characterization of lossless functions can be obtained via the well-known positive real lemma (e.g., [14]) which, for the case where the Hermitian part is to be identically zero, specializes to the following.

Proposition 3: A rational function $D + Cz(I_n - zA)^{-1}B$ belongs to $\mathbb{F}$ and has Hermitian part identically equal to zero a.e. on the boundary of the unit circle if and only if there exists $P \geq 0$ such that

$$P - A^*PA = 0 \tag{60}$$
$$C^* - A^*PB = 0 \tag{61}$$
$$D + D^* - B^*PB = 0. \tag{62}$$

Proof: The nonnegativity of

$$\begin{bmatrix} P - A^*PA & C^* - A^*PB \\ C - B^*PA & D + D^* - B^*PB \end{bmatrix} \tag{63}$$

along with $P \geq 0$ is equivalent to $D + Cz(I_n - zA)^{-1}B \in \mathbb{F}$ by the positive real lemma (see [14, p. 70]). Now, consider its Hermitian part

$$[B^*z^{-1}(I_n - z^{-1}A^*)^{-1} \quad I_m] \times \begin{bmatrix} O_n & C^* \\ C & D + D^* \end{bmatrix} \begin{bmatrix} z(I_n - z^{-1}A)^{-1}B \\ I_m \end{bmatrix}$$

and note that the null space of the mapping

$$M \mapsto \mathcal{G}(z)^* M \mathcal{G}(z)$$

where

$$\mathcal{G}(z) = \begin{bmatrix} z(I_n - zA)^{-1}B \\ I_m \end{bmatrix}$$

consists of matrices of the form

$$\begin{bmatrix} P - A^*PA & -A^*PB \\ -B^*PA & -B^*PB \end{bmatrix}.$$

It readily follows that if conditions (60)–(62) hold, then the function is lossless. If on the other hand (60)–(62) do not hold and (63) is simply nonnegative but not zero, then it can be shown that the Hermitian part can be factored into the product of nonzero spectral factors (cf. [14, p. 125]). ■

Returning to (57), in case A_o has all its eigenvalues in the open disc $\mathbb{D}$, then (48) is valid and (58) holds as well for all θ. In case A_o has eigenvalues on $\partial\mathbb{D}$, we need to decompose $F(z)$ into a lossless and a lossy summands. To do this, select T_1, T_2 matrices whose vectors form bases for the eignespaces of A corresponding to eigenvalues on $\partial\mathbb{D}$ and those in the interior of the disc, respectively. Then, A_o transforms into a block diagonal matrix

$$T^{-1}A_oT = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

where $T := [T_1, T_2]$ and the spectrum of A_1 is on the boundary and of A_2 in the interior of the unit disc, respectively. The input and output matrices B_o, C_o transform conformably into

$$T^{-1}B_o = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$
$$C_oT = [C_1 \quad C_2].$$

and

$$F(z) = D_\Psi + C_1 z(I - zA_1)^{-1}B_1 + C_2 z(I - zA_2)^{-1}B_2.$$

Then, we need to determine a value for a constant D_1 so that

$$F_1(z) = D_1 + C_1 z(I - zA_1)^{-1}B_1$$

is lossless. Necessarily, the remaining term $D_\Psi - D_1 + C_2 z(I - zA_2)^{-1}B_2$ is in $\mathbb{F}$ and is devoid of singularities on the boundary.

The transformation T_1 above, can be chosen so that A_1 is unitary, since A_1 has only simple eigenvalues on $\partial\mathbb{D}$. Then condition (60) leads to the commutation $A_1P = PA_1$ and, hence, that P is a polynomial function of A_1, i.e., that

$$P = p(A_1) := p_0 I + p_1 A_1 + \cdots + p_{n_1-1}A_1^{n_1-1}$$

The vector of coefficients $[p_0 \ \ldots \ p_{n_1-1}]$ can now be computed from (61) which becomes

$$A_1 C^* = p(A_1)B_1.$$

When $m > 1$, this is an overdetermined set of equations which is certain to have a solution. Finally, we may take

$$D_1 = \frac{1}{2}B^* p(A_1) B$$

to satisfy (62) and ensure that $F_1(z)$ is lossless. The matricial residues which represent the discontinuities in $d\mu(\theta)$ can now be computed by taking suitable limits at the singularities of A_1

$$V_\ell \rho_\ell V_\ell^* = \mathrm{Herm}\{\lim_{z\to e^{j\theta_\ell}}(1 - ze^{j\theta_\ell})F_1(z)\}, \ell = 1, 2, \ldots$$

Evidently, if A_1 is first brought into a diagonal form, then a convenient closed expression for the limit can be given in terms of partitions of B_1, C_1 corresponding to the eigenvalue $e^{j\theta_\ell}$.

IX. Impossibility of Decomposition Into White Noise + Deterministic Part

For the case of a scalar stochastic process $\{u_k : k \in \mathbb{Z}\}$, where $m = 1$, any state-covariance $\mathbf{R}$ can be written as

$$\mathbf{R} = \mathbf{R}_{\text{signal}} + \mathbf{R}_{\text{white noise}}$$

where

$$\mathbf{R}_{\text{white noise}} = \alpha_0 \mathbf{R}_0$$

with $\mathbf{R}_0$ being the solution to the Lyapunov equation

$$\mathbf{R}_0 - A\mathbf{R}_0 A^* = BB^*$$

and α_0 the smallest eigenvalue of the pencil $\mathbf{R} - \alpha\mathbf{R}_0$, i.e.,

$$\alpha_0 = \min\{\alpha : \det(\mathbf{R} - \alpha\mathbf{R}_0) = 0\} \quad (64)$$
$$= \max\{\alpha : \mathbf{R} - \alpha\mathbf{R}_0 \geq 0\}. \quad (65)$$

The matrix $\mathbf{R}_0$ is the controllability Grammian of the pair (A, B) and represents the state-covariance when the input is unit-variance white noise. Then $\mathbf{R}_{\text{white noise}}$ represents the maximal summand of $\mathbf{R}$ that can be attributed to a white-noise input component of (2), while the remaining $\mathbf{R}_{\text{signal}}$ corresponds to a deterministic input part. It can also be shown that this decomposition is canonical in the sense that any other one, consistent with a "white noise plus deterministic part" hypothesis for the input, will have a larger number of deterministic components (i.e., spectral lines). This is the interpretation of the CFP decomposition. The theory was originally developed for $\mathbf{R}$'s having a Toeplitz structure [26], [32] and extended to general state-covariances in [17] and [18].

It is rather instructive to present a derivation of the fact that, when $m = 1$, the equivalent conditions iii) and iii-a) of Proposition 1 are automatically satisfied by any singular state-covariance. This underscores the dichotomy with the multivariable case where a decomposition of $\mathbf{R}$ consistent with a "white noise plus deterministic part" input is not always possible (see Examples 1 and 2).

Proposition 4: Let $\mathbf{R}, A, B, H$ satisfy (5b) in Theorem 1, let $\mathbf{R} \geq 0$ and singular, and let $m = 1$. Then $B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B$ is invertible (i.e., $B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B$ is nonzero, since it is only a scalar quantity).

Proof: Suppose that $B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B$ is not invertible. Then

$$\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B = O_{n\times 1}, \quad \text{and} \quad (66)$$
$$\mathcal{R}(B) \subseteq \mathcal{R}(\mathbf{R}). \quad (67)$$

From (5b) and (66), it follows that $\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}A\mathbf{R}A^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})} = O_{n\times n}$, and hence, that $\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}A\mathbf{R} = O_{n\times n}$. From (67), $\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}AB = O_{n\times 1}$. By induction, using (5b), it follows that

$$\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}A^\ell\mathbf{R} = O_{n\times n}, \quad \text{for } \ell = 0, 1, \ldots$$

and hence, that $\mathcal{R}(\mathbf{R})$ is A-invariant. But $\mathcal{R}(B) \subseteq \mathcal{R}(\mathbf{R})$ and so is the largest A-invariant subspace containing $\mathcal{R}(B)$. Because (A, B) is a reachable pair, $\mathcal{R}(\mathbf{R}) = \mathbb{C}^n$ which contradicts the hypothesis that $\mathbf{R}$ is singular. ■

The following example shows that the statement of the proposition is only valid when $m = 1$ and that, in general, a decomposition of $\mathbf{R}$ consistent with a "white noise plus deterministic part" input is not always possible.

Example 2: Let

$$A = \begin{bmatrix} O_2 & O_2 \\ I_2 & O_2 \end{bmatrix}, \quad B = \begin{bmatrix} I_2 \\ O_2 \end{bmatrix}$$

and

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 1/2 & 3/4 \\ 0 & 1 & 0 & 1/2 \\ 1/2 & 0 & 1 & 0 \\ 3/4 & 1/2 & 0 & 1 \end{bmatrix}$$

where, as usual, I_2 and O_2 are the 2×2 identity and zero matrices, respectively. It can be readily seen that they satisfy conditions (4) as well as (5b) in Theorem 2—$\mathbf{R}$ being a block-Toeplitz matrix. Then $\mathbf{R}\geq 0$ and singular. To see this note that the first three principal minors of $\mathbf{R}$ are positive definite while

$$[-2 \quad -1 \quad 1 \quad 2]\mathbf{R} = O_{1\times 4}.$$

If the input to (2) is white noise with variance the 2×2 non-negative matrix

$$Q = \begin{bmatrix} a & b \\ \bar{b} & c \end{bmatrix}$$

then the state-covariance (for the chosen values of (A, B) and corresponding to this white-noise input) is

$$\mathbf{R}_0 = \begin{bmatrix} Q & O_2 \\ O_2 & Q \end{bmatrix} = I_2 \otimes Q.$$

We claim that

$$\mathbf{R} - \mathbf{R}_0 \geq 0 \Rightarrow \mathbf{R}_0 = O_{4\times 4}.$$

To prove this, consider that $Q\geq 0$ from which we obtain

$$ac \geq |b|^2, a\geq 0, c\geq 0. \tag{68}$$

Now, if $v = [-2 \quad -1 \ \ 1 \ \ 2]$ then $v\mathbf{R}v' = 0$ and $v\mathbf{R}_0 v' \geq 0$. Therefore

$$\begin{aligned}
&\mathbf{R} - \mathbf{R}_0 \geq 0 \\
&\quad\Rightarrow v\mathbf{R}_0 v' = 0 \\
&\quad\Rightarrow [2 \quad 1]Q\begin{bmatrix} 2 \\ 1 \end{bmatrix} + [1 \quad 2]Q\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 0 \\
&\quad\Rightarrow 5a + 4\Re e(b) + 5c = 0.
\end{aligned} \tag{69}$$

Thence, if $\beta := \Re e(b)$

$$\begin{aligned}
&ac \geq \beta^2 \\
&\quad\Rightarrow ac \geq \frac{25}{16}(a^2 + c^2 + 2ac) \\
&\quad\Rightarrow 0 \geq a^2 + c^2 + \frac{34}{25}ac \\
&\quad\Rightarrow \text{ either } a = 0 \text{ or } c = 0.
\end{aligned}$$

In either case, $|b| = 0$ and hence all three $a = b = c = 0$ from (69). Thus, $Q = O_{2\times 2}$ and $\mathbf{R}_0 = O_{4\times 4}$ as claimed. □

While the previous example shows that no white noise component can be subtracted in the hope of reaching a state-covariance satisfying condition iii) in Proposition 1 (thus corresonding to pure sinusoids), more is true. The following example shows that the off-diagonal block-entries of a block-Toeplitz $\mathbf{R}$ already prevent condition (iii) from being true.

Example 3: Let A, B as in Example 2 and

$$\mathbf{R} = \begin{bmatrix} a & b & 1/2 & 3/4 \\ \bar{b} & c & 0 & 0 \\ 1/2 & 0 & a & b \\ 3/4 & 0 & \bar{b} & c \end{bmatrix}.$$

In order for condition iii) of Proposition 1 to hold, the null space $\mathcal{N}(\mathbf{R})$ must have a dimension $\geq 2 = \dim(\mathcal{R}(B))$ (which can also readily seen from condition iii-a) as well). We argue that this cannot happen. Since

$$\begin{bmatrix} a & 3/4 \\ 3/4 & c \end{bmatrix}$$

is a principle minor of $\mathbf{R}\geq 0$, neither a nor c can vanish. The rank of

$$R_0 := \begin{bmatrix} a & b \\ \bar{b} & c \end{bmatrix}$$

must be equal to one, since there is a 3×3 minor of $\mathbf{R}$ with determinant $c\times\det(R_0)$. Hence, $ac = |b|^2 \Rightarrow c = |b|^2/2a$. But then, the northwest 3×3 principle minor of $\mathbf{R}$ is equal to $-c/4 < 0$, which contradicts $\mathbf{R}\geq 0$. □

X. Decomposition as a Convex Optimization Problem

We have just seen that in the case of a vectorial input, a decomposition of the state-covariance $\mathbf{R}$ of (2) which is consistent with the hypothesis of "white noise plus a deterministic signal at the input" may not always be possible (as demonstrated in, e.g., Example 3). Yet, the CFP dictum admits an alternative interpretation. Referring to (65) the CFP decomposition identifies the *maximal-variance* white noise component at the input. The fact that it leads to a splitting into white noise plus a deterministic signal for scalar inputs, may be seen as a lucky coincidence. Thus, for the case of multivariable inputs, we may interpret the CFP dictum as suggesting the selection of a maximal-variance white input component. In general, such a component will have a covariance Q which needs to be non-negative definite, i.e.,

$$Q \geq 0. \tag{70}$$

Then, also, the state covariance $\mathbf{R}_{\text{noise}}$ which is due to such a white noise input is the unique solution of the Lyapunov equation

$$\mathbf{R}_{\text{noise}} - A\mathbf{R}_{\text{noise}}A^* = BQB^*$$

and needs to satisfy

$$\mathbf{R}_{\text{noise}} \geq 0 \tag{72}$$

$$\mathbf{R}_{\text{signal}} := \mathbf{R} - \mathbf{R}_{\text{noise}} \geq 0. \tag{73}$$

Condition (72) follows from (71) together with the standing assumption (4). The remaining (70), (71), and (73) define a subset of $\mathbb{H}_m$ which is convex. Thus, the CFP dictum may be reinterpreted to seek a suitable extremal element in this set. In particular, we may choose to select a Q so as to account for a maximal-trace summand of $\mathbf{R}$ which may be due to a white noise input. This is formulated as follows.

Problem 1: Given $\mathbf{R}, A, B$ satisfying (4), $\mathbf{R} \geq 0$, and (5a) in Theorem 1, determine a decomposition

$$\mathbf{R} = \mathbf{R}_{\text{signal}} + \mathbf{R}_{\text{noise}} \tag{74}$$

where the summands satisfy (70), (71), (73) and $\mathbf{R}_{\text{noise}}$ has maximal trace, i.e.,

$$\mathbf{R}_{\text{noise}} = \operatorname{argmax}\{\ \operatorname{trace}(\mathbf{R}_{\text{noise}}) : (70), (71) \textit{and } (73)\ \text{hold}\}. \tag{75}$$

This is a standard convex optimization problem since the total noise variance $\operatorname{trace}(\mathbf{R}_{\text{noise}})$ is a linear functional of the parameters in Q and all constraints appear in the form of linear matrix inequalities. Thus, it can be readily and efficiently solved with existing computational tools [7].

It should be pointed out that, in general, there is no maximal element with respect to the partial order induced by non-negative definiteness and hence, alternatives to (75) corresponding to a different "normalizations" are also possible and give different solutions. For instance,

$$\mathbf{R}_{\text{noise}} = \operatorname{argmax}\{\ \operatorname{trace}(\mathbf{R}_{\text{noise}}\mathbf{W}) : (70),\ (71), \text{and } (73)\ \text{hold}\} \tag{76}$$

with $\mathbf{W} > 0$ a "weight" to encapsulate "prior" information about the directionality of the noise, or

$$Q = \operatorname{argmax}\{\ \operatorname{trace}(Q) : (70),\ (71),\ (73)\ \text{hold}\} \tag{77}$$

may be used instead.

Later, we present an example which shows that a maximum-trace solution as above, in general, does not lead to a decomposition with $\mathbf{R}_{\text{signal}}$ corresponding to a deterministic signal [i.e., satisfying (35)] even when an alternative decomposition does. Thus, the condition on invertibility of $B^*\mathbf{R}_{\text{signal}}B$ in Theorem 4 is not expected to hold in general and the theory in Section VIII may be used to construct the respective power spectra.

Example 4: With A, B as in Example 1, consider the state-covariance

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & 1/2 & 3/4 \\ r_1 & r_0 & 0 & 1/2 \\ 1/2 & 0 & r_0 & r_1 \\ 3/4 & 1/2 & r_1 & r_0 \end{bmatrix}$$

where the block-diagonal entries are yet unspecified. The values for these entries can be explicitly computed in the following two cases:

i) $B^*\mathbf{\Pi}_{\mathcal{N}(\mathbf{R})}B$ is invertible;
ii) $\operatorname{trace}(\mathbf{R})$ is minimal;

while always $\mathbf{R} \geq 0$.

The first can be carried out as follows. Condition i) is equivalent to the existence of a matrix

$$\Gamma = \begin{bmatrix} 1 & 0 & \gamma_{1,3} & \gamma_{1,4} \\ 0 & 1 & \gamma_{2,3} & \gamma_{2,4} \end{bmatrix}$$

such that $\Gamma\mathbf{R}$ is the zero matrix. Denote

$$R_0 := \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix}, \quad R_1 := \begin{bmatrix} 1/2 & 3/4 \\ 0 & 1/2 \end{bmatrix}, \quad \text{and} \quad \Gamma_0 := \begin{bmatrix} \gamma_{1,3} & \gamma_{1,4} \\ \gamma_{2,3} & \gamma_{2,4} \end{bmatrix}.$$

Since

$$R_0 + \Gamma_0 R_1^* = O_2 \Rightarrow R_0 + R_1\Gamma_0^* = O_2, \text{ while}$$
$$R_1 + \Gamma_0 R_0 = O_2$$

we deduce that

$$R_1 - \Gamma_0 R_1 \Gamma_0^* = O_2 \tag{78}$$
$$R_1 - \Gamma_0^2 R_1^* = O_2. \tag{79}$$

Equation (78) leads to $R_1 + R_1^* = \Gamma_0(R_1 + R_1^*)\Gamma_0^*$ and, if we factor $R_1 + R_1^* = SS^*$ with

$$S = \begin{bmatrix} 1 & 0 \\ 3/4 & \sqrt{1 - \left(\frac{3}{4}\right)^2} \end{bmatrix}$$

we deduce that $S^{-1}\Gamma_0 S$ must be unitary. Then from (79) we determine the eigenvalues of Γ_0. Carrying out all computations explicitly leads to

$$R_0 = \frac{1}{2}\begin{bmatrix} \frac{1}{\cos(\theta)} & \tan(\theta) \\ \tan(\theta) & \frac{1}{\cos(\theta)} \end{bmatrix}$$

and

$$\Gamma_0 = \begin{bmatrix} \frac{\cos(2\theta)}{\cos(\theta)} & -\tan(\theta) \\ \tan(\theta) & \frac{1}{\cos(\theta)} \end{bmatrix}$$

where $\sin(\theta) = (3/4)$. The values in R_0 is the unique set of values for which i) holds.

Similarly, the computation of the state-covariance with minimal trace as in ii) can be carried out explicitly to give

$$R_{0,\text{min trace}} = \begin{bmatrix} 3/4 & 1/2 \\ 1/2 & 3/4 \end{bmatrix}.$$

Finally, it is easy to check that $R_0 - R_{0,\text{min trace}}$ is indefinite.□

XI. Short-Range Correlation Structure

The rationale for the CFP decomposition has been re-cast in Problem 1 as seeking to extract the maximal variance that can be attributed to white-noise. In the case where $\mathbf{R}$ is block-Toeplitz as in (1), this amounts to determining a block-diagonal matrix

$\mathbf{R}_{\rm noise}$ of maximal trace satisfying the required positivity constraints (70), (71), (73). Yet, it is rarely the case in practice that a "white-noise" hypothesis is valid. Thus, we herein propose a new paradigm-a paradigm that also leads to a convex optimization problem and encompasses the above interpretation of the CFP decomposition as a special case. We seek to identify a maximal-variance summand which has a "short-range correlation structure" defined as follows.

Definition 1: Given A, B satisfying (4) a state-covariance $\mathbf{R}$ of the system (2) has *correlation range* k if there exists a matrix $H \in \mathbb{C}^{m\times n}$ so that

$$H^* = [B \quad AB \quad \dots \quad A^kB]\begin{bmatrix} Q_0^* \\ Q_1^* \\ \vdots \\ Q_k^* \end{bmatrix} \tag{80}$$

for suitable matrices $Q_0, \dots, Q_k$, such that

$$\mathbf{R} - A\mathbf{R}A^* = BH + H^*B^* \tag{81}$$

and

$$Q_0 + 2zQ_1 + \dots + 2z^kQ_k \in \mathbb{F}. \tag{82}$$

It is insightful to first consider the case where A, B are given as in (3) and the state-covariance structure is $(\ell+1)\times(\ell+1)$ block-Toeplitz. A block-Toeplitz matrix $\mathbf{R}$ has correlation range k if it is block-banded with all entries beyond the kth one being zero and, most importantly, it remains a covariance matrix when extended with zero elements beyond the ℓth entry as well. This is equivalent to $R_{k+1} = R_{k+2} = \dots = R_\ell = \dots = O_m$ being an admissible extension since already

$$R_0 + 2zR_1 + \dots + 2z^kR_k \in \mathbb{F}$$

from (82) because $Q_i = R_i$ $(i = 1, \dots, k)$ and $R_0 = Q_0^* + Q_0$.

Example 5: The following elementary example helps illustrate the concept of bounded correlation range. Consider the Toeplitz matrix

$$\mathbf{R} = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1 & 1/2 \\ 1/3 & 1/2 & 1 \end{bmatrix}.$$

We seek a Toeplitz noise-covariance summand of maximal trace with correlation range 1, i.e., we seek

$$\mathbf{R}_{\rm noise} = \begin{bmatrix} q_0 & q_1 & 0 \\ q_1 & q_0 & q_1 \\ 0 & q_1 & q_0 \end{bmatrix}$$

so that $\mathbf{R} - \mathbf{R}_{\rm noise} \geq 0$, and $q_0 + 2zq_1 \in \mathbb{F}$. Since $q_0 + 2zq_1$ is only of degree one, $q_0 + 2zq_1 \in \mathbb{F}$ if and only if $|q_1| \leq (1/2)q_0$. The solution turns out to be $q_0 = 2/3$ and $q_1 = 0.3097$.

Instead, if we sought $\mathbf{R}_{\rm noise}$ diagonal corresponding to white noise, the answer would have been $\mathbf{R}_{\rm noise} = \min\{\mathrm{eig}(\mathbf{R})\} \times I_3$. It can be easily checked that $\min\{\mathrm{eig}(\mathbf{R})\} = 0.4402 < q_0$. Thus, colored MA-noise allows a larger amount of energy to be accounted for. □

Problem 1 with condition (71) replaced by

$$\mathbf{R}_{\rm noise} \text{ having correlation range } k \tag{83}$$

is also a convex optimization problem. In general, the positive-real constraint (82) can be expressed as a convex condition via the well-known positive-real lemma (e.g., see [14]), and the maximizer of the trace can be readily obtained with existing numerical tools (e.g., the Matlab LMI toolbox).

In the case (2) has nontrivial dynamics, the right hand side of (81) becomes

$$\begin{aligned} BH + H^*B^* = {} & A^kBQ_k^*B^* + \dots + ABQ_1^*B^* \\ & + B(Q_0^* + Q_0)B^* + BQ_1B^*A^* + \dots + BQ_kB^*(A^*)^k \end{aligned}$$

and $\mathbf{R}$ can be interpreted as the state covariance due to colored noise at the input with spectral density

$$\begin{aligned} Q_k^*e^{-j\theta} + \dots + Q_1^*e^{-j\theta} + (Q_0^* + Q_0) \\ + Q_1e^{j\theta} + \dots + Q_ke^{jk\theta}. \end{aligned}$$

In a follow-up publication [22] we explain the relevance of such a decomposition for Toeplitz $\mathbf{R}$ and, accordingly, scalar input and band-Toeplitz $\mathbf{R}_{\rm noise}$ as in Example 5. A detailed study on the full potential of decomposition according to "correlation range" for high resolution spectral analysis will be presented in a forthcoming report.

XII. Concluding Remarks

The CFP decomposition underlies many subspace identification techniques in modern spectral analysis (such as MUSIC, ESPRIT, and their variants [32]). But in spite of its importance and its extensive appearance in many guises in the identification and signal processing literature, no multivariable analog had been proposed. Perhaps the reason can be sought in the fact that the exact analog of the CFP-decomposition does not exist. This realization led us to alternative interpretations of the CFP-decomposition, and the goal of this paper has been to explore such alternatives for a "signal plus noise" decomposition of covariances for multivariable processes. In the process, we have found that (e.g., see Example 3 and Section IX) regardless of how much of the energy is accounted for by noise, the remaining energy, in general, cannot be accounted for by pure spectral lines only. The remaining energy necessarily corresponds to a singular covariance matrix and thus, Sections VII and VIII develop the needed theory to construct spectra for singular matrices. Finally Sections X and XI develop certain alternatives to the CFP decomposition where we forgo the requirement that one part is completely deterministic, and allow instead that it has a long range correlation structure.

Acknowledgment

The author would like to thank Dr. D. Herrick for his input and for discussions that partially motivated this work.

REFERENCES

[1] D. Z. Arov, "The generalized bitangent Carathéodory-Nevanlinna-Pick problem, and (j, J_0)-inner matrix-valued functions," *Russian Acad. Sci. Izv. Math.*, vol. 42, no. 1, pp. 1–26, 1994.

[2] J. A. Ball, I. C. Gohberg, and L. Rodman, *Interpolation of Rational Matrix Functions, Operator Theory: Advances and Applications*. Boston, MA: Birkhäuser, 1990, vol. 45.

[3] 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. Oper. Theory*, vol. 9, pp. 107–142, 1983.

[4] C. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for Nevanlinna-Pick interpolation: A convex optimization approach to certain problems in systems and control," *IEEE Trans. Autom. Control*, vol. 45, no. 6, pp. 822–839, Jun. 2001.

[5] C. Carathéodory and L. Fejér, "Über den Zusammenghang der Extemen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landauschen Satz," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, pp. 218–239, 1911.

[6] T. Constantinescu, *Schur Parameters, Dilation and Factorization Problems*. Boston, MA: Birkhäuser-Verlag, 1996, vol. 82, p. 253.

[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*. Philadelphia, PA: SIAM, Jun. 1994.

[8] P. Delsarte, Y. Genin, and Y. Kamp, "Orthogonal polynomial matrices on the unit circle," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 3, pp. 149–160, Mar. 1978.

[9] P. Delsarte, Y. Genin, and Y. Kamp, "Generalized Schur representation of matrix-valued functions," *SIAM J. Alg. Disc. Meth.*, vol. 2, no. 2, pp. 94–107, 1981.

[10] P. Delsarte, Y. Genin, and Y. Kamp, "Schur parameterization of positive definite block-Toeplitz systems," *SIAM J. Appl. Math.*, vol. 36, no. 1, pp. 34–46, 1979.

[11] P. Dewilde and H. Dym, "Lossless chain scattering matrices and optimum linear predition: The vector case," *Circuit Theory Appl.*, vol. 9, pp. 135–175, 1981.

[12] H. Dym, *Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces, and Interpolation*. Providence, RI: AMS, 1989, vol. 71.

[13] H. Dym and I. Gohberg, "A maximum entropy principle for contractive interpolants," *J. Functional Anal.*, vol. 65, pp. 83–125, 1986.

[14] P. Faurre, M. Clerget, and F. Germain, *Operateurs Rationnels Positifs*. Paris, France: Dunod, 1979.

[15] C. Foias and A. Frazho, "The commutant lifting approach to interpolation problems," in *Operator Theory: Advances and Applications*. Basel, Germany: Birkhäuser-Verlag, 1990, vol. 44.

[16] C. Foias, A. Frazho, I. Gohberg, and M. A. Kaashoek, "Metric Constrained Interpolation, Commutant Lifting and Systems," in *Operator Theory: Advances and Applications*. Basel, Germany: Birkhäuser-Verlag, 1991, vol. 100.

[17] T. T. Georgiou, "Signal estimation via selective harmonic amplification: MUSIC, redux," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 780–790, Mar. 2000.

[18] T. T. Georgiou, "Spectral estimation via selective harmonic amplification," *IEEE Trans. Autom. Control*, vol. 46, no. 1, pp. 29–42, Jan. 2001.

[19] ——, "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.

[20] ——, "Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parameterization," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1811–1823, Nov. 2002.

[21] ——, "Relative entropy and the multi-variable multi-dimensional Moment Problem," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2007.

[22] ——, "Decomposition of Toeplitz matrices via convex optimization," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 537–540, Sep. 2006, to appear.

[23] T. T. Georgiou and M. C. Smith, "Optimal robustness in the gap metric," *IEEE Trans. Autom. Control*, vol. 35, no. 6, pp. 673–686, Jun. 1990.

[24] Ya. L. Geronimus, *Orthogonal Polynomials*. New York: Springer-Verlag, 1961, p. 570.

[25] I. Gohberg, Ed., "I. Schur Methods in Operator Theory and Signal Processing," in *Operator Theory Adv. Appl.*. Boston, MA: Birkhäuser, 1986, vol. 18.

[26] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. New York: Chelsea, 1958.

[27] S. Haykin, *Nonlinear Methods of Spectral Analysis*. New York: Springer-Verlag, 1979, p. 247.

[28] T. Kato, *Perturbation Theory for Linear Operators*. New York: Springer-Verlag, 1966.

[29] M. G. Krein and A. A. Nudel'man, *The Markov Moment Problem and Extremal Problems*. Providence, RI: AMS, 1977, vol. 50, p. 417.

[30] P. Masani, "Recent trends in multivariate prediction theory," in *Multivariate Analysis*, P. R. Krishnaiah, Ed. New York: Academic, 1966, pp. 351–382.

[31] M. Rosenblum and J. Rovnyak, *Hardy Classes and Operator Theory*. New York: Oxford Univ. Press, 1985.

[32] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.

Tryphon T. Georgiou (F'00) was born in Athens, Greece, on October 18, 1956. He received the Diploma in mechanical and electrical engineering from the National Technical University of Athens, in 1979, and the Ph.D. degree from the University of Florida, Gainesville, in 1983.

He has served on the faculty of Florida Atlantic (1983–1986) and Iowa State (1986–1989) Universities. Since 1989, he has been with the University of Minnesota, Minneapolis, where he holds the Vincentine Hermes-Luh Chair of Electrical Engineering.

Dr. Georgiou has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the *SIAM Journal on Control and Optimization*, and the *Systems and Control Letters*. He has served on the Board of Governors of the IEEE Control Systems Society. He has received the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society three times, in 1992, 1999, and 2003. In 1992 and 1999, he received the award for joint work with Prof. M. C. Smith (Cambridge Univ., U.K.), and in 2003 for joint work with Prof. C. Byrnes (Washington Univ., St. Louis, MO) and Prof. A. Lindquist (KTH, Stockholm, Sweden).