Relative Entropy and the Multivariable Multidimensional Moment Problem

Tryphon T. Georgiou, *Fellow, IEEE*

Abstract—Entropy-like functionals on operator algebras have been studied since the pioneering work of von Neumann, Umegaki, Lindblad, and Lieb. The best known are the von Neumann entropy $\mathbb{I}(\rho) := -\mathrm{trace}(\rho \log \rho)$ and a generalization of the Kullback–Leibler distance $\mathbb{S}(\rho \,\|\, \sigma) := \mathrm{trace}(\rho \log \rho - \rho \log \sigma)$, referred to as quantum relative entropy and used to quantify distance between states of a quantum system. The purpose of this paper is to explore $\mathbb{I}$ and $\mathbb{S}$ as regularizing functionals in seeking solutions to multivariable and multidimensional moment problems. It will be shown that extrema can be effectively constructed via a suitable homotopy. The homotopy approach leads naturally to a further generalization and a description of all the solutions to such moment problems. This is accomplished by a renormalization of a Riemannian metric induced by entropy functionals. As an application, we discuss the inverse problem of describing power spectra which are consistent with second-order statistics, which has been the main motivation behind the present work.

Index Terms—Covariance realization, moment problem, multivariable, multidimensional, quantum entropy, spectral analysis.

I. INTRODUCTION

THE quantum relative entropy (Umegaki [78])

$$\mathbb{S}(\rho \,\|\, \sigma) := \mathrm{trace}(\rho \log \rho - \rho \log \sigma)$$

where ρ, σ are positive Hermitian matrices (or operators) with trace equal to one, generalizes the Kullback–Leibler relative entropy [52], just as the von Neumann entropy [64]

$$\mathbb{I}(\rho) := -\mathrm{trace}(\rho \log \rho)$$

generalizes the classical Shannon entropy. They both inherit a rather rich structure from their scalar counterparts and in particular, $\mathbb{S}(\cdot \,\|\, \cdot)$ is jointly convex in its arguments as shown by Lieb [57] in 1973, whereas $\mathbb{I}(\cdot)$ is concave. The relative entropy originates in the quest to quantify the difficulty in discriminating between probability distributions and can be thought as a distance between such. Its matricial counterpart $\mathbb{S}$ can similarly be used to quantify distances between positive matrices.

Entropy and relative entropy have played a central rôle in thermodynamics in enumerating states consistent with data and, thereby, used to identify "the most likely" ones among all possible alternatives. The measurement of a physical property in a classical setting is modeled via *ensemble averaging* (e.g., see [30, Ch. 3])

$$r = \sum_k g(k)\rho(k)$$

where k runs over all microstates corresponding to a scalar value $g(k)$. Each microstate occurs with probability $\rho(k)$ and r is a moment of the underlying probability distribution. Similarly, quantum measurement is also modeled by averaging (as originally idealized by von Neumann, see, e.g., [77, Ch. 5], [43, p. 183])

$$\rho_{\mathrm{after}} = \sum_k G(k)\rho_{\mathrm{before}}G(k)^*$$

where the ρ's represent density matrices (positive Hermitian with trace one), the G's represent products of projection operators, and "$*$" denotes "conjugate-transpose." Similar expressions arise for the density operator when restricted to a subsystem (partial trace [77, p. 185]) and also when measuring "non-self-adjoint observables" (e.g., [81]). If the underlying Hilbert space is infinite dimensional then the measurement process can be modeled with a continuous analogue of the above where the summation is replaced by an integral (e.g., see [8]). These are instances of moment problems. More generally we may consider

$$R = \sum_k G_{\mathrm{left}}(k)\rho(k)G_{\mathrm{right}}(k) \tag{1}$$

where $\rho(k)$ are Hermitian positive matrices as well as its "continuous" counterpart

$$R = \int_{\mathcal{S}} G_{\mathrm{left}}(\theta)\rho(\theta)G_{\mathrm{right}}(\theta)\,d\theta \tag{2}$$

where $\rho(\theta)$ represents a Hermitian-valued positive (density) function on a support set $\mathcal{S} \subseteq \mathbb{R}^k (k \geq 1)$ and $G_{\mathrm{left}}, G_{\mathrm{right}}$ are matrix-valued functions on $\mathcal{S}$. If the underlying distribution is not absolutely continuous then we write

$$R = \int_{\mathcal{S}} G_{\mathrm{left}}(\theta)d\mu(\theta)G_{\mathrm{right}}(\theta)$$

instead, where $d\mu$ is such a positive Hermitian-valued measure.

The moment problem (1), (2) is typified by spectral analysis based on second-order statistics, especially in the context of sensor arrays and of polarimetric radar. The echo at different polarizations and/or at different wavelengths is being sampled at a variety of sensor locations. It is usually the case that these

samples are not independent and that the echo at different frequencies, polarizations, etc., affects each sensor by a different amount. Attributes of the scattering field (e.g., reflectivity at different wavelengths and polarization) and the relative position of the array elements with respect to the scattering field are responsible for the variations in the vectorial echo. The vector of attributes can be thought of as a vectorial input $u(\theta)$ to the array while the relative position and characteristics of its elements specify an $n_{\text{left}} \times m$ transfer function matrix

$$G_{\text{left}}(\theta) = \begin{bmatrix} g_{1,\text{left}} \\ \vdots \\ g_{n_{\text{left}},\text{left}} \end{bmatrix}$$

to the n_{left} sensor outputs. If the attributes $u(\theta)$ are modeled as a zero-mean vectorial stochastic process, independent over frequencies, then

$$y_{\text{left}} = \int_{\mathcal{S}} G_{\text{left}}(\theta) \, du(\theta)$$

represents the vectorial output process. Similarly, if

$$G_{\text{right}}(\theta) = \begin{bmatrix} g_{1,\text{right}}, & \cdots & g_{n_{\text{right}},\text{right}} \end{bmatrix}$$

is the $m \times n_{\text{right}}$ complex conjugate transpose of the transfer matrix corresponding to a second group of sensors, and if

$$y_{\text{right}} = \int_{\mathcal{S}} G_{\text{right}}(\theta)^* du(\theta)$$

designate the corresponding vector of n_{right} outputs, then the $n_{\text{left}} \times n_{\text{right}}$ correlation matrix

$$R = E\{y_{\text{left}} y_{\text{right}}^*\}$$

gives rise to the matricial moment constraint on the spectral distribution of u given in (2). On the other hand, (1) can be interpreted when the power spectrum is discrete. A power density which matches the correlation samples aims at giving clues about the makeup of the scattering field.

We address the moment problem in the above generality and provide a way to answer the following:

i)   does there exist a density function satisfying (1) and (2)?
ii) if yes, describe all density functions consistent with (1) and (2).

In essence, the above questions go back to 1980, when Dickinson [21] raised the issue of consistency of two-dimensional Markovian estimates. Consistency of *scalar* distributions was taken up in the work of Lang and McClellan [53] and Lewis [56]. Both references used entropy functionals and suggested computational solutions to the first question when dealing with scalar distributions. The present work follows in the footsteps of these as well as of a rather extensive literature on inverse problems [18], [19], [80], [54], [61], [17], [27], [55], [48], [9], [46] having roots in the early days of statistical mechanics. The key idea has been to seek extrema of entropy functionals—existence would guarantee solvability of the moment problem. However, the idea of using "weighted" entropy functionals to parametrize solutions originates in Byrnes, Gusev, and Lindquist [14] in dealing with a moment problem with "degree constraints." It

was followed up in [10], [11] for more general kernels, and in [10], [38] where it was reformulated using the Kullback–Leibler distance between sought solutions and positive "priors."

Classical moment problems [1], [51] are closely related to analytic interpolation ones, and as such, have been studied in great generality, including their matrix-valued counterpart (see, e.g., [71]). However, analytic interpolation applies only when the integration kernels possess a very particular shift-structure similar to that of a Fourier vector (see, e.g., [4], [23] and also [33], [34]) and is of limited use in the generality sought herein. It was the recent interest in interpolation with a "complexity/degree constraint" which forced a departure from the groove of the classical theory—seeking a characterization of solutions of a certain degree. The "trigonometric moment problem with degree constraint" was studied in [28] in both the scalar and the matricial setting using degree and homotopy methods (see also [29], [31]), motivated by Kalman [49] who raised the question of determining interpolants of minimal degree. Subsequent literature on analytic interpolation with "degree/complexity constraint" dealt mostly with scalar problems (see, e.g., [32], [15], [25], [16], and the references therein), and a most pertinent development was the aforementioned use of minimizers of weighted entropy functionals [14], [10] to construct solutions for general kernels as well. Then Blomqvist *et al.* [6] studied matricial Nevanlinna–Pick interpolation with similar tools. The framework of the present work is similarly based on entropy functionals and continues along the lines of [38] where the possibility of using quantum relative entropies was brought up.

In view of a rather rich literature on quantum entropies, a comment is in order as to other possible connections to the present work. Besides the Umegaki–von Neumann entropy $\mathbb{S}(\cdot \| \cdot)$ studied in this paper, there is a plethora of alternatives due to a dichotomy between matricial and scalar distributions [67], [66]. In particular, Araki's theory [3], [58] helped characterize a family of "quasi-entropies," contractive under stochastic maps. References [68], [69], [58] in particular explore the Riemannian geometry they induce on density matrices. It is an interesting question as to which among this "garden of entropies," besides the Umegaki–von Neumann one, allows a convenient representation of solutions for general moment problems. The approach we have taken leads us to work mostly with an induced metric (a Jacobian related to the Hessian of $\mathbb{S}(\cdot \| \cdot)$). A suitable normalization then recovers any solution of the moment problem as a corresponding extremal. It is not known whether, the "weighted metrics," e.g., ∇h_σ in Section IV, are metrics induced by a quasi-entropy in the language of Petz [68].

Finally, it is interesting to point out that, a counterpart for discrete distributions relates to the theory of analytic centers in semidefinite programming [7], [63]. In fact, a key construction in this paper—a homotopy for the numerical computation of solutions, is analogous to tracing paths of analytic centers in interior point methods.

In Section II, we introduce five motivating examples. Section III develops the geometry of matricial cones and the significance of relative entropy as a barrier functional. In order to simplify the exposition, we first deal with cones of constant matrices and then with those of matrix-valued density functions.

Except for technical differences, the two run in parallel. Finally, Section IV discusses the parametrization of solutions to the moment problem, followed by concluding remarks (Section V) and an Appendix with useful facts on matricial calculus.

II. Motivating Examples

In this section, we present five examples that motivate our study. For simplicity, the first two are developed in the context of scalar distributions. The last three are intrinsically multivariable. In particular, the fourth example pertains to the connection of the moment problem with analytic interpolation, and the fifth relates to approximation by positive matrices. We return to the first, fourth, and fifth example again later on in the paper.

A. Nonequispaced Arrays

Consider an array of sensors with three elements, linearly spaced at distances 1 and $\sqrt{2}$ wavelengths from one another, and assume that (monochromatic) planar waves, originating from afar, impinge upon the array. This is exemplified in Fig. 1.

Assuming that the sensors are sensitive to disturbances originating over one side of the array, with sensitivity independent of direction, the signal at the ℓth sensor is typically represented as a superposition

$$u_\ell(t) = \int_0^\pi A(\theta) e^{j(\omega t - p x_\ell \cos(\theta) + \phi(\theta))} d\theta$$

of waves arising from all spatial directions $\theta \in [0, \pi]$, where ω is as usual the angular time–frequency (as opposed to "spatial"), x_ℓ the distance between the ℓth and the 0th sensor, p the wavenumber, and $A(\theta)$ the amplitude and $\phi(\theta)$ a random phase of the θ-component. Typically, $\phi(\theta)$ for various values of θ are uncorrelated. The term $p x_\ell \cos(\theta)$ in the exponent accounts for the phase difference between reception at different sensors. For simplicity, we assume that $p = 1$ in appropriate units. Correlating the sensor outputs we obtain

$$R_k = E\{u_{\ell_1} \bar{u}_{\ell_2}\} := \int_0^\pi e^{-jk\cos(\theta)} f(\theta) d\theta$$

where $f(\theta) = |A(\theta)|^2$ represents *power density*, and $k = \ell_1 - \ell_2$ with $\ell_1 \geq \ell_2$ and belonging to $\{0, 1, \sqrt{2} + 1\}$. Thus,

$$k \in \mathcal{I} := \{0, 1, \sqrt{2}, \sqrt{2} + 1\}. \tag{3}$$

The only significance of our selection of distances between sensors, that gave rise to the indexing set (3), is to underscore that there is no algebraic dependence between the elements of the so-called *array manifold*

$$G(\theta) := \begin{bmatrix} 1 & e^{-j\tau} & e^{-j\sqrt{2}\tau} & e^{-j(\sqrt{2}+1)\tau} \end{bmatrix}'$$

(where "$[\cdot]'$" denotes the transpose, G is thought of as a column vector, and $\tau = \cos(\theta) \in [-1, 1]$).

Given a set of values R_k for $k \in \mathcal{I}$, it is often important to determine whether they are indeed the moments of a power density $f(\theta)$, and if so to characterize all consistent power spectra. The case of arrays with equispaced elements is very special and answers to such questions relate to the nonnegativity of a Toeplitz
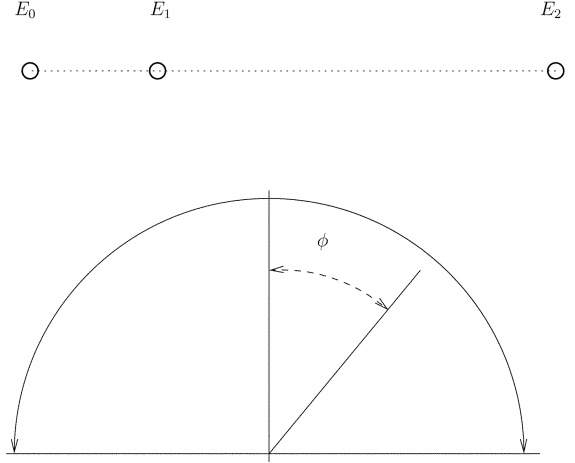

Fig. 1. Nonequispaced sensor array.

matrix formed out of the R_k's. In the present situation nonnegativity of

$$\int_{-1}^{1} \begin{bmatrix} 1 \\ e^{-j\tau} \\ e^{-j\sqrt{2}\tau} \end{bmatrix} \frac{f(\cos^{-1}(\tau))}{\sqrt{1-\tau^2}} \begin{bmatrix} 1 & e^{j\tau} & e^{j\sqrt{2}\tau} \end{bmatrix} d\tau$$

which, in the obvious indexing turns out to be

$$\begin{bmatrix} R_0 & R_1 & R_{\sqrt{2}+1} \\ \bar{R}_1 & R_0 & R_{\sqrt{2}} \\ \bar{R}_{\sqrt{2}+1} & \bar{R}_{\sqrt{2}} & R_0 \end{bmatrix}, \tag{4}$$

is only a necessary condition. The fact that it is *not* sufficient (see e.g., [30, p. 786]) motivated the present study.

B. Two-Dimensional Distributions

The subject matter of multidimensional distributions received considerable attention in the 1970s and early 1980s (e.g., see [59]). However, despite the rich theory of analytic interpolation and orthogonal polynomials in more than one variable, etc. (e.g., see [45], [73]), as pointed out by Dickinson [21], the analog of the questions raised earlier never received a definitive answer.

For instance, consider the simplest possible planar grid

$$\mathcal{I} := \{k, \ell \in \mathbb{Z} : 0 \leq k \leq n, 0 \leq \ell \leq n\}$$

which is both regular and square. In fact, we may even assume that $n = 2$. Then, consider the case where (monochromatic as before) waves impinge upon sensors at the grid nodes, the sensors are sensitive to reception on one side of their plane, the waves are planar, originating from all possible directions in the sky and uncorrelated over different directions, and that correlations at the sensor outputs are taken. The array manifold in this case becomes (after suitable normalization assumptions) the 3×3 array

$$G(\theta, \phi) = \begin{bmatrix} e^{j(k\theta + \ell\phi)} \end{bmatrix}_{k,\ell=0}^{k,\ell=2}$$

with $(\theta, \phi) \in [0, \pi] \times [0, \pi]$. This is quite standard (e.g., [47], [79], [42]). Now, if $\rho(\theta, \phi)$ denotes a positive scalar power density for the waves originating from (normalized) Euler angles

θ, ϕ then, correlation between the sensor outputs gives us a 3×3 matrix of covariance samples

$$\left[R_{k,\ell} := \int_0^\pi \int_0^\pi e^{j(k\theta+\ell\phi)} \rho(\theta,\phi) d\theta d\phi \right]_{k,\ell=0}^{k,\ell=2}.$$

The same two questions that we raised earlier are again relevant: Given a 3×3 array, how can we tell that it originates as suggested above, and how can we characterize *all* power densities that are consistent with the covariance samples $R_{k,\ell}$?

Lang and McClellan [53] considered "maximum entropy spectra" and suggested evaluating those on a discrete grid.[1] Lewis [56] considered a framework which is applicable for answering the existence question, while [35] discussed parametrization of solutions as well. The present contributions allows addressing the most general situation where the sensor array elements receive vectorial echoes and hence, $R_{k,\ell}$ as well as $\rho(\theta,\phi)$ are matrices.

C. Quantum Measurements

We temporarily adopt the language of quantum mechanics as in, e.g., [65], and explain how this relates to the linear algebraic framework in the Introduction. We then discuss a basic academic paradigm which exemplifies the setting of the matricial moment problem.

Let ρ^{AB} denote the density matrix of a quantum system composed of two subsystems A and B. Each subsystem can be in two states $|0\rangle$ and $|1\rangle$, respectively. These can be thought of as the vectors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{in } \mathbb{C}^2.$$

Then, the states of the combined system, $|i\rangle \otimes |j\rangle$ where $i, j \in \{0, 1\}$, can be represented by a vector in $\mathbb{C}^4$. For instance, $|0\rangle \otimes |1\rangle$ corresponds to

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

where in the last equation $\otimes$ is the Kronecker tensor product. Accordingly, ρ^{AB} is a Hermitian 4×4 trace one matrix formed out of sums of Kronecker products of 2×2 density matrices of the two subsystem. Now, if ρ^A represents the density matrix of subsystem A, then this can be obtained from ρ^{AB} via taking the partial trace with respect to subsystem B (e.g., see [65, pp. 106–107]). An important example is that of the Bell state with

$$\rho^{AB} = \left(\frac{|01\rangle + |11\rangle}{\sqrt{2}} \right) \left(\frac{\langle 00| + \langle 11|}{\sqrt{2}} \right) = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

The partial trace then gives (see [65, pp. 106–107])

$$\rho^A = \text{trace}_B(\rho^{AB}) = \frac{1}{2} I_2$$

where I_2 is the 2×2 identity matrix. If ρ^{AB} is represented as a 4×4 matrix as above, then

$$\rho^A = \sum_{k=1}^2 G_k \rho^{AB} G_k^*$$

where

$$G_1 = I_2 \otimes (1 \quad 0) = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and, similarly, $G_2 = I_2 \otimes (0 \quad 1)$. The reverse task of characterizing all possible ρ^{AB} based on a known ρ^A is a (discrete) moment problem.

D. State Statistics and Analytic Interpolation With Degree Constraint

Consider the linear discrete-time state equations

$$x_k = Ax_{k-1} + Bu_k, \qquad \text{for } k \in \mathbb{Z} \tag{5}$$

where $x_k \in \mathbb{C}^n, u_k \in \mathbb{C}^m, A \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{n \times m}, (A, B)$ is a controllable pair, and the eigenvalues of A lie in the open unit disk of the complex plane. Let $\{u_k : k \in \mathbb{Z}\}$ be a zero-mean stationary stochastic process with power spectrum the nonnegative matrix-valued measure $d\mu(\theta)$ on $\theta \in (-\pi, \pi]$. Then, *under stationarity conditions*, the state covariance

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

can be expressed in the form of the integral (cf. [60, Ch. 6])

$$R = \int_{-\pi}^\pi \left(G(e^{j\theta}) \frac{d\mu(\theta)}{2\pi} G(e^{j\theta})^* \right) \tag{6}$$

where

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

is the transfer function of system (5). Note that we use z to denote the transform of the delay operator and therefore $G(z)$ is analytic in the unit disc of the complex plane.

It turns out that state covariances R are characterized by the following two equivalent conditions (see [33], [34]):

$$\text{rank} \begin{bmatrix} R - ARA^* & B \\ B^* & 0 \end{bmatrix} = 2m \tag{7}$$

and

$$R - ARA^* = BH + H^*B^*, \quad \text{for some } H \in \mathbb{C}^{m \times n}. \tag{8}$$

Then, power spectral measures consistent with (6) are in correspondence with matrix-valued functions $F(z)$ on the unit circle $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ which have nonnegative real part via the Herglotz representation

$$F(z) = \int_{-\pi}^\pi \left(\frac{1 + ze^{j\theta}}{1 - ze^{j\theta}} \right) \frac{d\mu(\theta)}{2\pi} + jc \tag{9}$$

---

[1]Lang and McClellan [53] were the first to point out that existence cannot be guaranteed for the usual "entropy rate" functional $\int_{\mathcal{S}} \log(\rho) d\theta$ unless the dimension of $\mathcal{S}$ is one, cf. Theorem 4 below.

with jc an arbitrary skew-Hermitian constant. The measure $d\mu$ can be recovered as the weak* limit of the real part of $F(z)$ as z tends to the boundary, i.e.,

$$d\mu(\theta) \sim \lim_{r \nearrow 1} \Re(F(re^{j\theta})). \tag{10}$$

The class of nonnegative real-matrix-valued functions F giving rise to admissible power spectral measures are also characterized by the interpolation condition ([33])

$$F(z) = H(I - zA)^{-1}B + Q(z)V(z) \tag{11}$$

where Q is a matrix function analytic in $\mathbb{D}$,

$$V(z) := D + zC(I - zA)^{-1}B \tag{12}$$

and $C \in \mathbb{C}^{m \times n}, D \in \mathbb{C}^{m \times m}$ are selected so that V is inner, i.e., $V(\xi)^*V(\xi) = V(\xi)V(\xi)^* = I$ for all $|\xi| = 1$.

The data A, B, H, and $V(z)$ in (11) specify an analytic interpolation problem. Positive-real solutions to (11) can be given via (9) and solutions to the moment problem (6). The characterization of solutions to (6) given in Theorem 6 below, allows a nonclassical characterization of solutions to (11) and in particular a class of solutions of McMillan degree less than or equal to the dimension of (5) (see Section IV-B).

E. Conforming Sample State-Statistics To Underlying Dynamics

Consider the setting of Section II-D and note that the dynamics of the filter (5) impose a structural constraint on the state covariance matrix R. This is expressed in the form of the two equivalent conditions (7) and (8). For instance, in the case where u_k is scalar and

$$G(z) = \begin{bmatrix} 1 & z & \dots & z^n \end{bmatrix}',$$

representing a "delay line," the state covariance R in (6) has a Toeplitz structure.

When the state covariance is estimated from a short observation record $\{x_0, x_1, \dots, x_N\}$, the sample covariance

$$\hat{R} = \frac{1}{N+1} \sum_{k=0}^{N} x_k x_k^*$$

almost certainly will fail to be consistent with (7) and (8). Attention to this issue has been raised before and various options exist for the case where R is a Toeplitz matrix (see, e.g., the algorithm due to Burg in [41]). Yet, in general, it is natural to begin with the positive-definite sample covariance $\hat{R}$ and then seek a value for R which is positive definite, consistent with (7) and (8), and closest to $\hat{R}$ in a suitable sense.

For general matrix norms, respective optimization may be awkward due to the presence of the positivity constraint on R. An alternative is to use a von Neumann-like relative entropy as a distance measure between R and $\hat{R}$ (cf. [36], [37]). On the other hand, consistency with (8) can be expressed as moment constraints. This will be taken up again in Section IV-C after we develop solutions to general moment problems.

III. MATRICIAL DISTRIBUTIONS AND THEIR MOMENTS

The moment conditions (1) and (2) are linear constraints on densities $\rho_k, k = 1, 2, \dots$ and $\rho(\theta), \theta \in \mathcal{S}$, respectively. Density functions, whether discrete or continuous, are nonnegative, or nonnegative definite in the matricial case, for each value of their indexing set. Thus, they have the structure of a cone. En-

tropy functionals, on the other hand, represent natural barriers on such positive cones and can be used to identify, and even parametrize, density functions which are consistent with given moment conditions. We begin by explaining the geometry of the moment problem for constant density matrices and the relevance of entropy functionals in obtaining solutions as their respective extrema. Both the geometry of cones of matricial densities functions as well as the rôle of entropy functionals is quite similar and is taken up in Section III-B.

A. Relative Entropy and the Geometry of Matricial Cones

We begin by focusing on constraints

$$R = \sum_k G_{\text{left}}(k)\rho\, G_{\text{right}}(k)$$

where ρ is not indexed. The general case is quite similar.

We use the notation

$$\mathbb{M} := \{M \in \mathbb{C}^{m \times m} : M = M^*\}$$
$$\mathfrak{M} := \{M \in \mathbb{M} \text{ and } M \geq 0\}$$
$$\mathfrak{M}_+ := \{M \in \mathbb{M} \text{ and } M > 0\}$$

to denote the space of Hermitian matrices and the cones of nonnegative and positive-definite ones, respectively. The space $\mathbb{M}$ is endowed with a natural inner product

$$\langle M_1, M_2 \rangle := \text{trace}(M_1^* M_2) = \text{trace}(M_1 M_2)$$

as a linear space over $\mathbb{R}$. Clearly, both, $\mathfrak{M}$ and $\mathfrak{M}_+$ are convex cones. Since nonnegativity of $\langle M_1, M \rangle$ for all $M_1 \in \mathfrak{M}$ implies that $M \in \mathfrak{M}$, it follows that $\mathfrak{M}$ is self-dual.[2] It can also be seen that $\mathfrak{M}_+$ is the interior of $\mathfrak{M}$.

The linear operator

$$L : \mathbb{M} \to \mathfrak{R} : \rho \mapsto R = \sum_k G_{\text{left}}(k)\rho\, G_{\text{right}}(k)$$

where $\mathfrak{R} \subseteq \mathbb{C}^{n_{\text{left}} \times n_{\text{right}}}$ denotes the range of L, maps $\mathfrak{M}$ onto the cone of admissible moments $\mathcal{K} = L(\mathfrak{M}) \subseteq \mathfrak{R}$. Here, and throughout, $G_{\text{left}}, G_{\text{right}}$ are matrices of dimension $n_{\text{left}} \times m$ and $m \times n_{\text{right}}$, respectively. A further assumption that is often needed is that the null spaces of L and $\mathfrak{M}$ have no common point other than the origin, i.e.,

$$\text{null}(L) \cap \mathfrak{M} = \{0\}. \tag{13}$$

The interior of $\mathcal{K}$ is $\text{int}(\mathcal{K}) = L(\mathfrak{M}_+)$ and, given R, the moment problem requires testing whether $R \in \mathcal{K}$ and if so, characterizing all $\rho \in \mathfrak{M}$ such that $R = L(\rho)$.

Geometry in the range space $\mathfrak{R}$ is based on

$$\langle \lambda, R \rangle := \Re e(\text{trace}(\lambda^* R)), \qquad \text{for } \lambda, R \in \mathfrak{R}. \tag{14}$$

Then the adjoint transformation of L is

$$L^* : \mathfrak{R} \to \mathbb{M} : \lambda \mapsto \rho = \left(\sum_k G_{\text{left}}(k)^* \lambda\, G_{\text{right}}(k)^* \right)_{\text{Herm}}$$

where $(M)_{\text{Herm}} := \frac{1}{2}(M + M^*)$ is the "Hermitian part." The dual cone of $\mathcal{K}$

$$\mathcal{K}^{\text{dual}} := \{\lambda \in \mathfrak{R} : \langle \lambda, R \rangle \geq 0, \forall R \in \mathcal{K}\}$$

is naturally related to the cone $\mathfrak{M} \subset \mathbb{M}$. In fact, using $\langle \lambda, L(\rho) \rangle = \langle L^*(\lambda), \rho \rangle$, it follows easily that

$$\mathcal{K}^{\text{dual}} = \{\lambda \in \mathfrak{R} : L^*(\lambda) \in \mathfrak{M}\}.$$

---

[2]In general, the dual cone $\mathfrak{M}^{\text{dual}}$ is the set of elements forming an "acute angle" with all elements of the original cone, i.e., $\{M : \langle M, M_1 \rangle \geq 0, \forall M_1 \in \mathfrak{M}\}$ (see [51]).

The interior of the dual cone

$$\mathcal{K}_+^{\text{dual}} := \text{int}(\mathcal{K}^{\text{dual}}) := \{\lambda : \langle \lambda, R \rangle > 0, \ \forall R \in \mathcal{K} - \{0\}\}$$

corresponds to $\mathfrak{M}_+$ as is easily seen to satisfy

$$\mathcal{K}_+^{\text{dual}} = \{\lambda : L^*(\lambda) \in \mathfrak{M}_+\}.$$

Finally, (13) can be seen to be equivalent to $\mathcal{K}_+^{\text{dual}} \neq \emptyset$.

1) Minimizers of $\mathbb{S}(I \| \rho)$*:* We are interested in minimizers of (the negative entropy)

$$\mathbb{S}(I \| \rho) = -\text{trace}(\log(\rho))$$

on $\mathfrak{M}_{R,+} := \{\rho \in \mathfrak{M}_+ : R = L(\rho)\}$. Here and throughout, "$I$" denotes the identity matrix of size determined from the context. When such a minimizer exists at an interior point of $\mathfrak{M}_{R,+}$, stationarity conditions for the entropy functional dictate an explicit form for the minimizer (which, is unique due to the convexity of $-\text{trace}(\log(\rho))$).

The Lagrangian of the problem is

$$\mathcal{L}(\lambda, \rho) := \text{trace}(-\log(\rho)) - \langle \lambda, R - L(\rho) \rangle.$$

Using the expression for the derivative of the logarithm given in Proposition 8 of the Appendix, the (Gateaux) derivative of $\mathcal{L}$ in the direction $\delta \in \mathbb{M}$ becomes

$$d\mathcal{L}(\lambda, \rho; \delta) := \text{trace}(-M_\rho^{-1}\delta) + \langle \lambda, L(\delta) \rangle$$
$$= \text{trace}(-\rho^{-1}\delta) + \langle L^*(\lambda), \delta \rangle.$$

In the preceding derivation, the "trace" is what allows replacing the "noncommutative division operator" M_ρ^{-1} (cf. (48)) with multiplication by ρ^{-1}. The stationarity condition $d\mathcal{L}(\lambda, \rho; \delta) \equiv 0$ then gives

$$\rho = (L^*(\lambda))^{-1}. \tag{15}$$

Thus, a necessary condition is that there exist $\lambda \in \mathcal{K}^{\text{dual}}$ such that $L^*(\lambda)$ is strictly positive, i.e., that $\mathcal{K}_+^{\text{dual}}$ is nonempty. It turns out that if $R \in \text{int}(\mathcal{K})$ then this condition is also sufficient.

Theorem 1: Assume that $R \in \text{int}(\mathcal{K})$. Then the entropy functional $\mathbb{S}(I \| \rho)$ has a minimum in $\mathfrak{M}_{R,+}$, which is also unique, if and only if $\mathcal{K}_+^{\text{dual}}$ is nonempty.

Proof: Necessity is obvious and was argued above. Uniqueness follows from the matrix convexity of $\log(\cdot)$ (which follows, e.g., from a positive Hessian (47) given in the Appendix). Sufficiency requires that there exists $\lambda_1 \in \mathcal{K}_+^{\text{dual}}$ such that $R = L((L^*(\lambda_1))^{-1})$. Then $\rho_1 = (L^*(\lambda_1))^{-1}$ satisfies both the stationarity conditions and the contstraints. We show this via a continuity argument which we will adopt again later on to more general cases.

Consider the mapping

$$h : \mathcal{K}_+^{\text{dual}} \to \text{int}(\mathcal{K}) \subset \mathfrak{R}$$
$$: \lambda \mapsto L((L^*(\lambda))^{-1}).$$

Its Jacobian

$$\nabla h |_\lambda : \mathfrak{R} \to \mathfrak{R} : \delta \mapsto L(L^*(\lambda)^{-1} L^*(\delta) L^*(\lambda)^{-1})$$

is Hermitian since

$$\langle \delta_1, \nabla h(\delta) \rangle = \langle L^*(\delta_1), L^*(\lambda)^{-1} L^*(\delta) L^*(\lambda)^{-1} \rangle$$
$$= \langle L^*(\lambda)^{-1} L^*(\delta_1) L^*(\lambda)^{-1}, L^*(\delta) \rangle$$
$$= \langle \nabla h(\delta_1), \delta \rangle.$$

It is also invertible. To see this note that $\langle \delta, \nabla h(\delta) \rangle = 0$ implies that $L^*(\delta) = 0$, and since $\delta \in \mathfrak{R} = \text{range}(L)$, it implies that $\delta = 0$. Hence, the map h is a local diffeomorphism and ∇h can be used to relate locally a smooth path in the space of R's to one in the space of λ's (of course, both spaces being the same space $\mathfrak{R}$).

Choose a $\lambda_0 \in \mathcal{K}_+^{\text{dual}}$, let

$$\rho_0 = (L^*(\lambda_0))^{-1}$$

and let $R_0 = L(\rho_0)$. Consider the interval path $R_\tau = (1-\tau)R_0 + \tau R$, for $\tau \in [0, 1]$. Since $R \in \text{int}(\mathcal{K})$, so is the whole path R_τ ($\tau \in [0, 1]$). We claim that for all $\tau \in [0, 1]$ there exists $\lambda_\tau \in \mathcal{K}_+^{\text{dual}}$ such that $R_\tau = L(L^*(\lambda_\tau)^{-1})$. It is clear that this holds locally and that λ_τ satisfies

$$\frac{d}{d\tau}\lambda_\tau = (\nabla h |_{\lambda_\tau})^{-1}(R - R_0) \tag{16}$$

since $\frac{d}{d\tau}R_\tau = R - R_0$. The starting point is λ_0 and (16) can be integrated over a maximal interval $[0, \epsilon)$ for which $\lambda_\tau \in \mathcal{K}_+^{\text{dual}}$. Throughout

$$R_\tau = L(L^*(\lambda_\tau)^{-1}).$$

If $\epsilon > 1$, this proves our claim. If $\epsilon \leq 1$, then either $\|\lambda_\tau\| \to \infty$ as $\tau \to \epsilon$, or the λ_τ's have a limit point λ_ϵ on the boundary of $\mathcal{K}_+^{\text{dual}}$, i.e., such that $L^*(\lambda_\epsilon)$ is singular. Below we argue that neither is possible, which then shows that $\epsilon > 1$ and completes the proof.

We first show that λ_τ remains bounded. Assume to the contrary, i.e., assume that $\|\lambda_\tau\|$ grows unbounded as $\tau \to \epsilon$, and let $\ell_\tau := \lambda_\tau / \|\lambda_\tau\|$ (where $\|\lambda\| = \sqrt{\langle \lambda, \lambda \rangle}$ as usual). Since $R_\epsilon \in \text{int}(\mathcal{K})$, it holds that $R_\epsilon = L(\rho)$ with $\rho > 0$, and $\langle \ell_\tau, R \rangle$ is bounded away from zero for elements $\ell_\tau \in \mathcal{K}^{\text{dual}}$ of unit norm. However, because

$$\langle \lambda_\tau, R_\tau \rangle = \langle L^*(\lambda_\tau), L^*(\lambda_\tau)^{-1} \rangle = \text{trace}(I)$$

it follows that $\langle \ell_\tau, R_\tau \rangle = \text{trace}(I)/\|\lambda_\tau\|$ has zero as a limit point when $\tau \in [0, \epsilon)$; hence, so does $\langle \ell_\tau, R_\epsilon \rangle$. But this is a contradiction, hence λ_τ remains bounded.

We finally show that $L^*(\lambda_\tau)^{-1}$ and ∇h along with its inverse remain bounded. Consider the quadratic form

$$\langle \delta, \nabla h_{\lambda_\tau}(\delta) \rangle = \|L^*(\lambda_\tau)^{-1/2} L^*(\delta) L^*(\lambda_\tau)^{-1/2}\|^2 \tag{17}$$

for $\delta \in \mathfrak{R}$. Since λ_τ (and, hence, $L^*(\lambda_\tau)$) remains bounded, the quadratic form is bounded away from zero when $\tau \in [0, \epsilon)$. Hence, $\nabla h |_{\lambda_\tau}^{-1}$ is uniformly bounded on $[0, \epsilon)$. On the other hand, because of (13), the minimal angle between any ray in the cone $\mathfrak{M}$ and $\text{range}(L^*)$ is bounded away from $\pi/2$. Hence, $\|R_\tau\| > \alpha\|L^*(\lambda_\tau)^{-1}\|$, for some $\alpha > 0$. But R_τ remains bounded. We conclude that $L^*(\lambda_\tau)^{-1}$ remains bounded and that ∇h remains bounded as well. This completes the proof. $\square$

2) Minimizers of $\mathbb{S}(\rho \,\|\, I)$: We now focus on minimizers of

$$\mathbb{S}(\rho \,\|\, I) = \mathrm{trace}(\rho \log(\rho))$$

in $\mathfrak{M}_+$, subject to $R = L(\rho)$. The Lagrangian this time is

$$\mathcal{L}(\lambda, \rho) := \mathrm{trace}(\rho \log(\rho)) - \langle \lambda, R - L(\rho) \rangle.$$

Once again, using the expression for the differential of the logarithm given in Proposition 8 of the Appendix, the (Gateaux) derivative of $\mathcal{L}$ in the direction $\delta \in \mathbb{M}$ becomes

$$d\mathcal{L}(\lambda, \rho; \delta) := \mathrm{trace}\left(\delta \log(\rho) + \rho M_\rho^{-1}(\delta)\right) + \langle \lambda, L(\delta) \rangle$$
$$= \mathrm{trace}(\delta \log(\rho) + \delta) + \langle L^*(\lambda), \delta \rangle.$$

The last step follows from

$$\mathrm{trace}\left(\rho M_\rho^{-1}(\delta)\right)$$
$$= \mathrm{trace}\left(\rho \int_0^\infty (\rho + tI)^{-1}\delta(\rho + tI)^{-1}dt\right)$$
$$= \mathrm{trace}\left(\int_0^\infty (\rho + tI)^{-1}\rho(\rho + tI)^{-1}dt\,\delta\right) = \mathrm{trace}(\delta).$$

The stationarity condition $d\mathcal{L}(\lambda, \rho; \delta) \equiv 0$ then gives that

$$\rho = \exp(-I - L^*(\lambda)) = \frac{1}{e}\exp(-L^*(\lambda)) \qquad (18)$$

with $L^*(\lambda) \in \mathbb{M}$ (and not necessarily in $\mathfrak{M}$ as before). It turns out that if $R \in \mathrm{int}(\mathcal{K})$ a minimizer can always be found. It should be noted that (13) is no longer a necessary condition.

Theorem 2: If $R \in \mathrm{int}(\mathcal{K})$, then the entropy functional $\mathbb{S}(\rho \,\|\, I)$ has a minimum in $\mathfrak{M}_{R,+}$ which is unique and of the form (18).

Proof: We use a similar continuity argument as before. Consider the mapping

$$\kappa : \mathfrak{R} \to \mathrm{int}(\mathcal{K}) \subset \mathfrak{R}$$
$$: \lambda \mapsto L\left(\frac{1}{e}\exp(-L^*(\lambda))\right). \qquad (19)$$

Its Jacobian

$$\nabla k\,|_\lambda : \mathfrak{R} \to \mathfrak{R} : \delta \mapsto \frac{1}{e}L(M_{\exp(-L^*(\lambda))}(-L^*(\delta))) \qquad (20)$$

with M_C as in (46), is Hermitian and negative definite. This is because

$$\langle \delta_1, \nabla\kappa\,|_\lambda(\delta) \rangle$$
$$= -\frac{1}{e}\langle L^*(\delta_1), M_{\exp(-L^*(\lambda))}(L^*(\delta)) \rangle$$
$$= -\frac{1}{e}\mathrm{trace}\left(L^*(\delta_1)\int_0^1 e^{-(1-t)L^*(\lambda)}L^*(\delta)e^{-tL^*(\lambda)}dt\right)$$
$$= -\frac{1}{e}\mathrm{trace}\left(\int_0^1 e^{-tL^*(\lambda)}L^*(\delta_1)e^{-(1-t)L^*(\lambda)}\,dt\,L^*(\delta)\right)$$
$$= \langle \nabla\kappa\,|_\lambda(\delta_1), \delta \rangle$$

while

$$\langle \delta, \nabla\kappa\,|_\lambda(\delta) \rangle$$
$$= -\frac{1}{e}\int_0^1 \mathrm{trace}\left(A^{\frac{t}{2}}L^*(\delta)A^{-\frac{t}{2}}AA^{-\frac{t}{2}}L^*(\delta)A^{\frac{t}{2}}\right)dt$$

with $A = \exp(-L^*(\lambda)) \in \mathfrak{M}_+$. Then $\langle \delta, \nabla\kappa\,|_\lambda(\delta) \rangle$ is clearly negative unless $\delta = 0$. Thus, the map κ is a local diffeomorphism and $\nabla\kappa$ can be used to relate locally a smooth path in the space of R's to one in the space of λ's.

Begin with λ_0 in $\mathfrak{R}$, $R_0 = L(\frac{1}{e}\exp(-L^*(\lambda_0)))$, and $R_\tau = (1 - \tau)R_0 + \tau R$ for $\tau \in [0, 1]$. Since $R_0, R \in \mathrm{int}(\mathcal{K})$ then $R_\tau \in \mathrm{int}(\mathcal{K})$ for all $\tau \in [0, 1]$. We claim that

$$\frac{d}{d\tau}\lambda_\tau = (\nabla\kappa\,|_{\lambda_\tau})^{-1}(R - R_0) \qquad (21)$$

can be integrated over $[0, 1]$ with λ_τ staying bounded. Then, by construction, λ_τ satisfies both

$$R_\tau = L\left(\frac{1}{e}\exp(-L^*(\lambda_\tau))\right) \qquad (22)$$

as well as the stationarity conditions for each τ. Hence,

$$\rho = \frac{1}{e}\exp(-L^*(\lambda_1))$$

is a minimizer for $\mathbb{S}(\rho \,\|\, I)$ as claimed in the proposition.

Clearly, (21) can be integrated over $[0, \epsilon)$. If $\epsilon > 1$, we are done. Thus, we only need to show that $\epsilon \leq 1$ and

$$\|\lambda_\tau\| \to \infty \text{ as } \tau \to \epsilon$$

lead to a contradiction. To this end, we use two facts, first that (22) holds on $[0, \epsilon)$ and then, that $\langle \ell, R_\epsilon \rangle > 0$ for all $\ell \in \mathcal{K}^{\mathrm{dual}}$, since $R_\epsilon \in \mathrm{int}(\mathcal{K})$.

Define $\ell_\tau = \lambda_\tau/\|\lambda_\tau\|$ and let ℓ_ϵ be a limit point of a convergent subsequence ℓ_{τ_i} with $\tau_i \to \epsilon$ (which exists since the ℓ_τ's are bounded). We claim that $L^*(\ell_\epsilon) \geq 0$ and singular. To see this first note that, if

$$\mathrm{spectrum}(L^*(\ell_\epsilon)) \cap (-\infty, 0) \neq \emptyset$$

then $\mathrm{spectrum}(L^*(\ell_{\tau_i})) \cap (-\infty, 0) \neq \emptyset$ as well, for sufficiently large i. But if this is so, then

$$L\left(\frac{1}{e}e^{-L^*(\lambda_{\tau_i})}\right) = L\left(\frac{1}{e}e^{-L^*(\ell_{\tau_i})\|\lambda_{\tau_i}\|}\right) \qquad (23)$$

grows without bound instead of tending to R_ϵ. Therefore,

$$\mathrm{spectrum}(L^*(\ell_\epsilon)) \subset [0, \infty).$$

Now, if $L^*(\ell_\epsilon)$ is nonsingular then for all i sufficiently large $L^*(\ell_{\tau_i})$ is nonsingular as well. But then, (23) tends to zero as $i \to \infty$. Hence, $0 \in \mathrm{spectrum}(L^*(\ell_\epsilon)) \subset [0, \infty)$.

Now let U be an isometry ($U^*U = I$) whose columns span the range of $L^*(\ell_\epsilon)$ and consider

$$\langle \ell_\epsilon, R_{\tau_i} \rangle = \frac{1}{e}\left\langle L^*(\ell_\epsilon), e^{-L^*(\lambda_{\tau_i})}\right\rangle$$
$$= \frac{1}{e}\left\langle L^*(\ell_\epsilon), e^{-L^*(\ell_{\tau_i})\|\lambda_{\tau_i}\|}\right\rangle$$
$$= \frac{1}{e}\left\langle U^*L^*(\ell_\epsilon)U, e^{-U^*L^*(\ell_{\tau_i})U\|\lambda_{\tau_i}\|}\right\rangle.$$

Since

$$U^*L^*(\ell_{\tau_i})U \to U^*L^*(\ell_\epsilon)U > 0$$

while $\|\lambda_{\tau_i}\| \to \infty$, we conclude that $\langle \ell_\epsilon, R_{\tau_i} \rangle \to 0$ as $i \to \infty$. Since $R_{\tau_i} \to R_\epsilon$, it follows that $\langle \ell_\epsilon, R_\epsilon \rangle = 0$ which contradicts the hypothesis that $R_\epsilon \in \mathrm{int}(\mathcal{K})$. $\qquad \square$

B. Relative Entropy and Matricial Distributions

The geometry of convex cones and of the moment problem when ρ is a matricial density function on a compact set $\mathcal{S}$, as in (1) and (2), is quite similar to the case where ρ is only a positive matrix as in Section III-A. Appropriate generalizations of the relative entropy functionals allow computable expressions for the corresponding extrema when $\mathcal{S}$ is a closed interval of the real line, or even a multidimensional closed interval in $\mathbb{R}^k (k > 1)$. We develop this theory focusing on (2).

We consider Hermitian $m \times m$ matrix-valued measurable functions on $\mathcal{S}$ as a linear space over $\mathbb{R}$ with an inner product

$$\langle m_1, m_2 \rangle = \int_{\mathcal{S}} \text{trace}(m_1(\theta) m_2(\theta)) \, d\theta.$$

We use the notation $\mathbb{M}$ to denote the Hilbert space of square integrable elements, and the notation $\mathcal{M}$ and $\mathcal{M}_+$ to denote the cones of elements which are nonnegative and positive definite, respectively, for all $\theta \in \mathcal{S}$. The linear operator

$$L : \mathcal{M} \to \mathfrak{R} : \rho \mapsto R = \int_{\mathcal{S}} G_{\text{left}}(\theta) \rho(\theta) G_{\text{right}}(\theta) \, d\theta \quad (24)$$

maps $\mathcal{M}$ into a subspace of $\mathbb{C}^{\text{left} \times \text{right}}$ denoted by $\mathfrak{R}$ as before and viewed as a linear space over $\mathbb{R}$. Both, moments R and their duals λ reside in $\mathfrak{R}$ and the geometry is always based on (14). For simplicity of the exposition, we assume that the integration kernels $G_{\text{left}}, G_{\text{right}}$ are *continuously differentiable* on $\mathcal{S}$. The *closure* of the range of $\mathcal{M}$ is denoted by $\mathcal{K} = \overline{L(\mathcal{M})}$, while $\text{int}(\mathcal{K}) = L(\mathcal{M}_+)$. The adjoint transformation is now

$$L^* : \mathfrak{R} \to \mathbb{M} : \lambda \mapsto \rho = (G_{\text{left}}(\theta)^* \lambda \, G_{\text{right}}(\theta)^*)_{\text{Herm}}.$$

It is not difficult to show that the expressions for the dual cone and its interior

$$\mathcal{K}^{\text{dual}} = \{\lambda \in \mathfrak{R} : L^*(\lambda) \in \mathcal{M}\} \quad \text{and}$$
$$\mathcal{K}_+^{\text{dual}} = \{\lambda \in \mathfrak{R} : L^*(\lambda) \in \mathcal{M}_+\}$$

remain valid (except for the obvious change where $\mathcal{M}$ replaces our earlier $\mathfrak{M}$). The analog of (13) will be needed (in Theorem 4) which, can also be expressed as

$$\mathcal{K}_+^{\text{dual}} \neq \emptyset. \quad (25)$$

Finally, we define as before

$$\mathcal{M}_{R,+} := \mathcal{M}_+ \cap \{\rho \in \mathbb{M} : R = L(\rho)\}$$

as we seek to determine whether or not $\mathcal{M}_{R,+} = \emptyset$, or equivalently, whether $R \in \text{int}(\mathcal{K})$.

For future reference, we bring in a characterization of elements $R \in \mathcal{K}$ analogous to the scalar real case given in [51, p. 14]. Given $R \in \mathfrak{R}$, define the real-valued functional

$$\mathfrak{C}_R : \mathfrak{R} \to \mathbb{R}$$
$$: \lambda \mapsto \langle \lambda, R \rangle. \quad (26)$$

Such a bounded functional is said to be *nonnegative* (resp., *positive*)—denoted by $\mathfrak{C}_R \geq 0$ (resp., $\mathfrak{C}_R > 0$), if and only if the infimum of $\mathfrak{C}_R(\lambda)$ over $\lambda \in \mathcal{K}_+^{\text{dual}}$ of unit norm is positive (resp., nonnegative).

Proposition 1: The following hold:

$$R \in \mathcal{K} \Leftrightarrow \mathfrak{C}_R \geq 0$$
$$R \in \text{int}(\mathcal{K}) \Leftrightarrow \mathfrak{C}_R > 0.$$

Proof: We now only prove necessity, which is needed in the Proof of Theorem 4. The proof of sufficiency will be given at the end Section III-B.1.

If $R \in \text{int}(\mathcal{K})$, there exists a particular $\rho \in \mathcal{M}_+$ such that $R = L(\rho)$. It readily follows that $\mathfrak{C}_R > 0$. If $R \in \mathcal{K}$, there exist an approximating sequence $R_i \to R \ (i = 1, 2, \ldots)$ where $R_i = L(\rho_i)$ and $\rho_i \in \mathcal{M}$. If $\lambda \in \mathcal{K}_+^{\text{dual}}$ then $\langle \lambda, R_i \rangle > 0$ and hence,

$$\mathfrak{C}_R(\lambda) = \lim_{i \to \infty} \langle \lambda, R_i \rangle \geq 0. \qquad \square$$

We now turn to relative entropy functionals for matricial distributions. Given $\rho, \sigma \in \mathcal{M}_+$

$$\mathbf{S}(\rho \| \sigma) := \int_{\mathcal{S}} \text{trace}(\rho \log \rho - \rho \log \sigma) d\theta. \quad (27)$$

Once again, minimizers of relative entropy subject to the moment constraints (2) take a particularly simple form amenable to a numerical solution via continuation methods. We follow the same plan as in Section III-A by focusing successively on each of the two alternative choices, $\mathbf{S}(I \| \rho)$ and then $\mathbf{S}(\rho \| I)$. A significant departure from the case of constant densities shows up when considering the dimension of the support set $\mathcal{S}$ in the context of $\mathbf{S}(I \| \rho)$.

1) Minimizers of $\mathbf{S}(I \| \rho) = -\int_{\mathcal{S}} \text{trace}(\log(\rho)) d\theta$: In complete analogy with constant case the derivative of the Lagrangian

$$\mathcal{L}(\lambda, \rho) := -\int_{\mathcal{S}} \text{trace}(\log(\rho)) d\theta - \langle \lambda, R - L(\rho) \rangle$$

in the direction $\delta \in \mathbb{M}$ is

$$d\mathcal{L}(\lambda, \rho; \delta) := \text{trace} \int_{\mathcal{S}} (-M_{\rho(\theta)}^{-1} + L^*(\lambda)) \delta(\theta) d\theta$$
$$= \text{trace} \int_{\mathcal{S}} (-\rho(\theta)^{-1} + L^*(\lambda)) \delta(\theta) \, d\theta$$

where, once again, the presence of the trace allows replacing the "super-operator" $M_{\rho(\theta)}^{-1}$ by multiplication by $\rho(\theta)^{-1}$, pointwise over $\mathcal{S}$. The fundamental lemma in calculus of variations (see, e.g., [5]) now gives the stationarity condition

$$\rho = L^*(\lambda)^{-1}. \quad (28)$$

In order for $\rho \in \mathcal{M}$, it is necessary that $L^*(\lambda)$ is strictly positive on $\mathcal{S}$. Thus, we consider the "rational" family of potential minimizers for $\mathbf{S}(I \| \rho)$

$$\mathcal{M}_{\text{rat}} := \{\rho = L^*(\lambda)^{-1}, \text{ with } \lambda \in \mathcal{K}_+^{\text{dual}}\}$$

where we seek a solution to the moment constraints (2). It turns out that if a solution exists then a particular one exists in $\mathcal{M}_{\text{rat}}$ and that it can be obtained by computing the fixed point of an exponentially converging matrix differential equation. This differential equation is an appropriate generalization of (16). We summarize all these conclusions as follows.

Theorem 4: If $\dim(\mathcal{S}) = 1$, condition (25) holds, and $R \in \text{int}(\mathcal{K})$, then $\mathbf{S}(I \| \rho)$ has a minimum in $\mathcal{M}_{R,+}$ which is unique

and belongs to $\boldsymbol{\mathcal{M}}_{\mathrm{rat}}$. Furthermore, for any $\lambda_0 \in \mathcal{K}_+^{\mathrm{dual}}$, the solution λ_t of the matrix differential equation

$$\frac{d}{dt}\lambda_t = (\nabla h\,|_{\lambda_t})^{-1}(R - L(L^*(\lambda_t)^{-1})) \qquad (29)$$

where

$$\nabla h\,|_{\lambda_t} : \mathfrak{R} \to \mathfrak{R} : \delta \mapsto L(L^*(\lambda_t)^{-1}L^*(\delta)L^*(\lambda_t)^{-1}) \quad (30)$$

belongs to $\mathcal{K}_+^{\mathrm{dual}}$ for all $t \in [0, \infty)$, it converges to a point $\hat{\lambda} \in \mathcal{K}_+^{\mathrm{dual}}$ as $t \to \infty$ corresponding to this unique minimizer $\rho = L^*(\hat{\lambda})^{-1}$ for $\mathbf{S}(I \,\|\, \rho)$ satisfying $R = L(\rho)$. The differential (29) is exponentially convergent as the square distance $V(\lambda_t) = \|R - L(L^*(\lambda_t)^{-1})\|^2$ satisfies

$$\frac{dV(\lambda_t)}{dt} = -2V(\lambda_t).$$

Conversely, if $R \notin \mathrm{int}(\mathcal{K})$ and the dimension of $\mathcal{S}$ is one, then the differential equation (29) diverges.

Equations (29) is equivalent to

$$\frac{d}{d\tau}\lambda_\tau = (\nabla h\,|_{\lambda_\tau})^{-1}(R - R_0) \qquad (31)$$

modulo scaling of the integration variable (see below). The latter can be integrated over $[0, 1]$, and then $\hat{\lambda} = \lambda_\tau\,|_{\tau=1}$, yet (29) appears preferable for numerical reasons.

We wish to point out that the assumption on the dimension of $\mathcal{S}$ can be slightly relaxed to being at most two provided $\mathcal{S}$ is a torus and $G_{\mathrm{left}}, G_{\mathrm{right}}$ doubly periodic accordingly, (cf. [53, Example 2 on p. 882], [35]). Also, the first part of the theorem, when specialized to the case of moment problems that relate to analytic interpolation ones, as in Section II-D, it generalizes the results in [6] who use a similar entropy minimization formalism (cf. Section IV-B and the concluding remarks).

Proof of Theorem 4: Once again we consider the mapping

$$h : \mathfrak{R} \to \mathfrak{R} : \lambda \mapsto R = L(L^*(\lambda)^{-1}).$$

The Jacobian $\nabla h\,|_\lambda$ is Hermitian and invertible when $\lambda \in \mathcal{K}_+^{\mathrm{dual}}$. The proof is identical to the one given in Section III-A.1. Choose $\lambda_0 \in \mathcal{K}_+^{\mathrm{dual}}$, set $R_0 = L(L^*(\lambda_0)^{-1})$, and consider the one-parameter homotopy of maps

$$L(L^*(\lambda_\tau)^{-1}) = R + \tau(R - R_0) =: R_\tau \qquad (32)$$

for $\tau \in [0, 1]$. The idea is to follow a path of solutions $\lambda_\tau (\tau \in [0, 1])$, ensure that this is contained in $\mathcal{K}_+^{\mathrm{dual}}$, and set $\hat{\lambda} = \lambda_1$ which then satisfies the sought conditions $(L^*(\hat{\lambda})^{-1} \in \boldsymbol{\mathcal{M}}_{\mathrm{rat}}$ and $L(L^*(\hat{\lambda})^{-1}) = R)$.

Clearly, $L^*(\lambda_0)^{-1} \in \boldsymbol{\mathcal{M}}_{\mathrm{rat}}$ and $R_0 \in \mathrm{int}(\mathcal{K})$. If $R \in \mathrm{int}(\mathcal{K})$ then so is R_τ for $\tau \in [0, 1]$. We claim that for all $\tau \in [0, 1]$ there exists $\lambda_\tau \in \mathcal{K}_+^{\mathrm{dual}}$ such that

$$R_\tau = L(L^*(\lambda_\tau)^{-1}) \qquad (33)$$

as before. The arguments are similar to those given in the Proof of Theorem 1 and are based on the fact that h is a local diffeomorphism. Values λ_τ which obey

$$\frac{d\lambda_\tau}{d\tau} = (\nabla h_{\lambda_\tau})^{-1}(R - R_0) \qquad (34)$$

satisfy (33) as long as the path stays in $\mathcal{K}_+^{\mathrm{dual}}$. We need to rule out λ_τ crossing the boundary of $\mathcal{K}_+^{\mathrm{dual}}$ or tending to ∞ at a $\tau \leq 1$. Either possibility contradicts $R_\tau \in \mathrm{int}(K)$ $(\tau \in [0, 1])$ in a way analogous to the earlier arguments in the Proof of Theorem 1.

We consider a maximal interval $[0, \epsilon)$ over which $R_\tau \in \mathrm{int}(K)$, and note that

$$\langle \lambda_\tau, R_\tau \rangle = \int_{\mathcal{S}} \mathrm{trace}(I)\,d\theta = \mathrm{trace}(I) \cdot \mathrm{measure}(\mathcal{S}) < \infty$$

on $[o, \epsilon)$. If $\|\lambda_\tau\| \to \infty$, then $\mathfrak{C}_{R_\tau}$ cannot be bounded away from zero since $\langle \frac{\lambda_\tau}{\|\lambda_\tau\|}, R_\tau \rangle \to 0$. Hence, $R_\epsilon \notin \mathrm{in}(\mathcal{K})$ and $\epsilon > 1$.

If λ_ϵ lies on the boundary of $\mathcal{K}^{\mathrm{dual}}$, then $L^*(\lambda_\epsilon)$ has a root in $\mathcal{S}$. It is here that the dimension of $\mathcal{S}$ becomes important. As long as the dimension of $\mathcal{S}$ is one, $L^*(\lambda_\epsilon)^{-1}$ is not integrable as a function of $\theta \in \mathcal{S}$ and $R_\epsilon = L(L^*(\lambda_\epsilon)^{-1})$ cannot be finite. Thus again, $\epsilon > 1$.

We finally express (34) in a "feedback form." We first replace τ with $t = -\log(1 - \tau)$. In this case, $\tau = 1 - e^{-t}$ and t varies in $[0, \infty)$ as τ varies in $[0, 1]$. If we denote $\lambda_t := \lambda_{\tau(t)}$ and $R_t := R_{\tau(t)}$, then

$$\begin{aligned}
\frac{d}{dt}\lambda_t &= \left(\frac{d\tau}{dt}\right)(\nabla h\,|_{\lambda_t})^{-1}(R - R_0) \\
&= (1 - \tau(t))(\nabla h\,|_{\lambda_t})^{-1}(R - R_0) \\
&= (\nabla h\,|_{\lambda_t})^{-1}(R - R_t).
\end{aligned} \qquad (35)$$

The same substitution gives $\frac{dR_t}{dt} = R_1 - R_t$, and that

$$V(\lambda_t) := \|R - R_t\|^2 = \|R - L(L^*(\lambda_t)^{-1})\|^2$$

satisfies

$$\begin{aligned}
\frac{dV(\lambda_t)}{dt} &= -2\left\langle R - L(L^*(\lambda_t)^{-1}), \nabla h\,|_{\lambda_t}\frac{d}{dt}\lambda_t \right\rangle \\
&= -2\langle R - R_t, R - R_t \rangle = -2V(\lambda_t).
\end{aligned}$$

Uniqueness of the representation $R = L(\rho)$ with $\rho \in \boldsymbol{\mathcal{M}}_{\mathrm{rat}}$ follows from the fact that such a ρ is a minimizer of the strictly convex functional $\mathbf{S}(I \,\|\, \rho)$ subject to the moment constraints. An alternative but equivalent argument can be based on the fact that h is a C^1-mapping between open convex subsets of a Euclidean space, with a positive definite Jacobian everywhere.

In the other direction, if $R \notin \mathrm{int}(\mathcal{K})$, then the path R_τ either crosses or at least, in case $R \in \partial\mathcal{K}$, tends to the boundary of $\mathcal{K}$. In either case, λ_t grows unbounded and the differential equation diverges. $\qquad \square$

We now complete the Proof of Proposition 3.

Proof ["Sufficiency" in Proposition 3]: In the Proof of Theorem 4, in essence, we used the positivity of $\mathfrak{C}_R$ to obtain a

representation $R = L(\rho)$ with $\rho \in \mathcal{M}_+$. Thus, the same line of argument gives that if $\mathfrak{C}_R > 0$ then $R \in \text{int}(\mathcal{K})$.

If $\mathfrak{C}_R \geq 0$ but not necessarily > 0, then chose $R_0 \in \text{int}(\mathcal{K})$ and note that $\mathfrak{C}_{R+\frac{1}{k}R_0} > 0$ for $k = 1, 2, \ldots$. Obviously, $R + \frac{1}{k}R_0 \in \text{int}(\mathcal{K})$ $(k \to \infty)$ and R is at least in $\mathcal{K}$.[3] $\qquad \square$

2) Minimizers of $\mathbf{S}(\rho \| I) = \int_{\mathcal{S}} \text{trace}(\rho \log(\rho)) d\theta$: Once again, the derivative of the Lagrangian

$$\mathcal{L}(\lambda, \rho) := -\int_{\mathcal{S}} \text{trace}(\rho \log(\rho)) d\theta - \langle \lambda, R - L(\rho) \rangle$$

in the direction $\delta \in \mathbb{M}$ is

$$d\mathcal{L}(\lambda, \rho; \delta) := \text{trace} \int_{\mathcal{S}} (\log(\rho) + I + L^*(\lambda)) \delta(\theta) \, d\theta.$$

The stationarity condition leads to the expression

$$\rho = \frac{1}{e} \exp(-L^*(\lambda)) \qquad (36)$$

for the minimizer, except that now ρ is a function of $\theta \in \mathcal{S}$. We consider the "exponential" family

$$\mathcal{M}_{\exp} := \left\{ \rho = \frac{1}{e} \exp(-L^*(\lambda)), \quad \text{with } \lambda \in \mathfrak{R} \right\}$$

of potential minimizers for $\mathbf{S}(\rho \| I)$, where we seek a solution to (2). The development runs in parallel to the case where $\rho \in \mathcal{M}_{\text{rat}}$ with one important difference. The "Lagrange multipliers" λ no longer need to be restricted to $\mathcal{K}^{\text{dual}}$ and existence of solutions when $R \in \mathcal{K}$ can be guaranteed even when $\dim(\mathcal{S}) > 1$. Moreover, (25) is no longer necessary and existence of a solution to the moment problem in $\mathcal{M}_{\exp}$ is impervious to the dimension of the dual cone $\mathcal{K}_+^{\text{dual}}$.

Theorem 5: If $R \in \text{int}(\mathcal{K})$ then the entropy functional $\mathbf{S}(\rho \| I)$ has a minimum in $\mathcal{M}_{R,+}$ which is unique and belongs to $\mathcal{M}_{\exp}$. Furthermore, for any $\lambda_0 \in \mathfrak{R}$, the solution λ_t of

$$\frac{d}{dt}\lambda_t = (\nabla k |_{\lambda_t})^{-1}(R - L(\lambda_t)) \qquad (37)$$

where

$$\nabla k |_{\lambda_t} : \mathfrak{R} \to \mathfrak{R} : \delta \mapsto -\frac{1}{e} L\left(\mathcal{M}_{\exp(-L^*(\lambda_t))}(L^*(\delta))\right) \quad (38)$$

remains bounded for $t \in [0, \infty)$ and converges to $\hat{\lambda} \in \mathfrak{R}$ as $t \to \infty$ corresponding to the unique minimizer $\rho = \frac{1}{e} \exp(-L^*(\hat{\lambda}))$ for $\mathbf{S}(I \| \rho)$ subject to $R = L(\rho)$. The convergence is exponential as $V(\lambda_t) = \|R - \frac{1}{e}L(\exp(-L^*(\hat{\lambda}_t)))\|^2$ satisfies

$$\frac{dV(\lambda_t)}{dt} = -2V(\lambda_t).$$

Conversely, if $R \notin \text{int}(\mathcal{K})$, then the differential (37) diverges.

Proof: The arguments are for the most part identical to those used in proving Theorem 2, i.e., we now consider

$$\kappa : \mathfrak{R} \to \text{int}(\mathcal{K}) \subset \mathfrak{R} : \lambda \mapsto R = L\left(\frac{1}{e}e^{-L^*(\lambda)}\right)$$

[3]A uniform bound on the integral of densities corresponding to $R + \frac{1}{k}R_0$ can be shown. This can be used to establish a finite nonnegative measure $d\mu$ corresponding to R.

observe that the Jacobian is negative definite for any value of λ, and use it to track a linear path $(1 - \tau)R_0 + \tau R$ ($\tau \in [0, 1]$) from $R_0 = L(\frac{1}{e}e^{-L^*(\lambda_0)})$ to the given R in the λ-coordinates. This is done by integrating (21) over $[0, 1]$ starting from arbitrarily chosen starting point λ_0. By construction, the solution of (21) corresponds to $\rho_\tau = \frac{1}{e}e^{-L^*(\lambda_\tau)} \in \mathcal{M}_+$ which satisfies $R_\tau = L(\rho_\tau)$. It is clear that if $R \notin \text{int}(\mathcal{K})$, then the differential equation diverges for $\tau \leq 1$ (since, otherwise, $\rho_\tau \in \mathcal{M}_+$ and $R_\tau = L(\rho_\tau)$ would hold on $[0, 1]$, contradicting $R \notin \text{int}(\mathcal{K})$). We only need to show that if $R \in \text{int}(\mathcal{K})$, then λ_τ remains bounded for $\tau \in [0, 1]$. This yields $\hat{\lambda} = \lambda_\tau |_{\tau=1}$ which satisfies $R = L(\frac{1}{e}e^{-L^*(\hat{\lambda})})$. Then (37) can be obtained via a change of variables as in Theorem 4. The same applies to deriving the differential equation for the "error" $V(\lambda_t)$.

In order to show that, in the event $R \in \text{int}(\mathcal{K})$, λ_τ remains bounded on $[0, 1]$ we extend the argument used to prove Theorem 2 to the present case where ρ is a matrix-valued function on $\mathcal{S}$. The key is to observe that, when (21) is integrated over a maximal interval $[0, \epsilon)$, any convergent subsequence of $\ell_\tau := \lambda_\tau / \|\lambda_\tau\|$ ($\tau \in [0, \epsilon)$) must have a limit point ℓ_ϵ for which $L^*(\ell_\epsilon) \in \mathcal{M}$ but not in $\mathcal{M}_+$. Moreover, $L^*(\ell_\epsilon)$ must be singular on $\mathcal{S}_0 \subset \mathcal{S}$ (a subset of possibly zero measure). To see this note the following. If $L^*(\ell_\epsilon) \in \mathcal{M}_+$, then $L^*(\ell_{\tau_i})$ is bounded away from zero and positive for i large enough, whereas if $L^*(\ell_\epsilon) \notin \mathcal{M}$ then there is a subset of $\mathcal{S}$ of nonzero measure where $L^*(\ell_{\tau_i})$ is negative. Either way

$$L(\exp(-L^*(\lambda_{\tau_i}))) = L(\exp(-L^*(\ell_{\tau_i} \cdot \|\lambda_{\tau_i}\|)))$$

cannot tend to R_ϵ as it should. In the first instance, it goes to zero and in the second it becomes unbounded. Thus, $L^*(\ell_\epsilon) \in \mathcal{M}$ but singular for certain values of θ. Below we show that this implies $R_\epsilon \notin \text{int}(\mathcal{K})$, which then proves that $\epsilon > 1$ and that (37) can be integrated on $[0, 1]$.

To show that $R_\epsilon \notin \text{int}(\mathcal{K})$ it suffices to show that $\mathfrak{C}_{R_\epsilon}$ is not strictly positive. To this end, we evaluate

$$\begin{aligned} \mathfrak{C}_{R_{\tau_i}}(\ell_\epsilon) &= \langle \ell_\epsilon, L(\exp(-L^*(\ell_{\tau_i}\|\lambda_{\tau_i}\|))) \rangle \\ &= \langle L^*(\ell_\epsilon), \exp(-L^*(\ell_{\tau_i}\|\lambda_{\tau_i}\|)) \rangle \\ &= \int_{\mathcal{S}} \text{trace}(L^*(\ell_\epsilon) \exp(-L^*(\ell_{\tau_i} \cdot \|\lambda_{\tau_i}\|))) d\theta \\ &= \int_{\mathcal{S}} \text{trace}(L^*(\ell_\epsilon)(\exp(-L^*(\ell_{\tau_i})))^{\|\lambda_{\tau_i}\|}) d\theta. \end{aligned}$$

For each value of θ, $(\exp(-L^*(\ell_{\tau_i})))^{\|\lambda_{\tau_i}\|}$ tends to zero outside the null space of $L^*(\ell_\epsilon)$. Since $\mathcal{S}$ is compact, the integrand goes to zero uniformly in θ as $i \to \infty$. Therefore, $\mathfrak{C}_{R_{\tau_i}}(\ell_\epsilon) \to 0$ as well, and $R \notin \text{int}(\mathcal{K})$. $\qquad \square$

C. Nonequispaced Arrays (Example II-A Continued)

We continue with Example II-A. We begin with a "true" density ρ_{true} shown in Fig. 2 and generate covariance samples R. This "true" density does not need to be in any particular form—computation of R is done via numerical integration.

Next, we integrate (29) and (37) taking $\lambda_0 = [1 \ 0 \ 0 \ 0]$, and display in Fig. 2 the resulting $\rho_{\exp}(\lambda_\infty, \theta)$ and $\rho_{\text{rat}}(\lambda_\infty, \theta)$, for comparison. Both are constructed using the fixed point of the
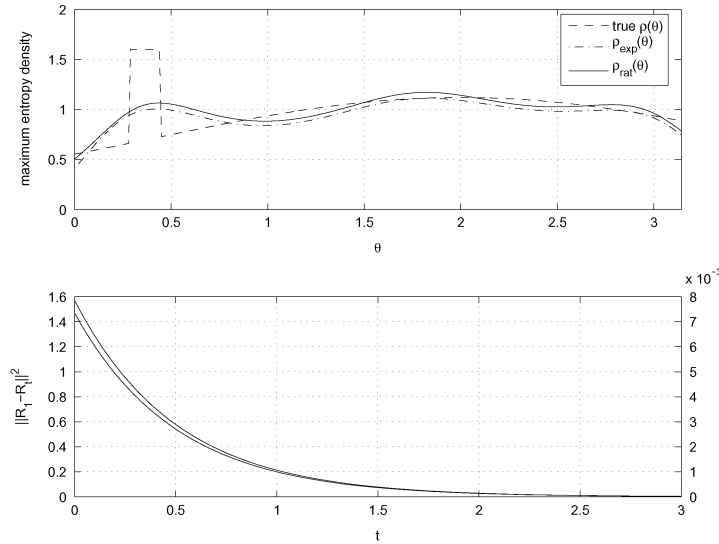
Fig. 2.   Computation of spectra for a non-equispaced array.

corresponding differential equations. The rate of convergence is the same, while the distance of the starting choice (for the same λ_0) may be different—as is the case here (with $\mathcal{M}_{\exp}$ corresponding to the y-axis to the left and $\mathcal{M}_{\mathrm{rat}}$ the labeling to the right in the subplot $(2,1)$).

IV. THE COMPLETE SET OF POSITIVE SOLUTIONS

Reference [38] suggested that all positive solutions to the moment problem may be obtained as minimizers of a suitable entropy functional, e.g., as being

$$\arg\min\{\mathbf{S}(\sigma\,\|\,\rho): R = L(\rho)\} \tag{39}$$

with σ thought of as a parameter. This was carried out successfully in [38], [10], [11], and [35] for the case where density functions are scalar-valued, for different levels of generality. Naturally, certain complications arise in the matricial setting. We discuss this next in the context of constant ρ, σ as in Section III-A. The generalization to the nonconstant case is straightforward and a positive result is given for the general case.

Considering the Lagrangian and the stationarity conditions for (39) we arive at

$$d\mathcal{L}(\lambda, \rho; \delta) = \mathrm{trace}(-\delta M_\rho^{-1}(\sigma)) + \langle L^*(\lambda), \delta\rangle$$
$$= \mathrm{trace}(-\delta M_\rho^{-1}(\sigma) + \delta L^*(\lambda))$$

leading to

$$M_\rho^{-1}(\sigma) = L^*(\lambda).$$

Although the "parameter" σ can be readily expressed as $M_\rho(L^*(\lambda))$, the density ρ which we are interested in, cannot be expressed in any effective way as a function of σ and the dual variable λ. Thus, a convenient functional form for the minimizer of (39) is unknown.

The option of minimizing $\mathbb{S}(\rho\,\|\,\sigma)$ subject to $R = L(\rho)$, however, goes through. Analysis of the corresponding Lagrangian readily leads to

$$\rho = \frac{1}{e}\exp(\log(\sigma) - L^*(\lambda)).$$

A computational theory, following the lines of Sections III-A.2 and III-B.2 easily carries through.

An attractive third alternative originates in the observation that the geometry of the problem, throughout, was inherited by the definiteness of the Jacobian maps. This suggests to forgo an explicit form for the entropy functional and start instead with a computable Jacobian. To this end we consider

$$h_\sigma : \lambda \mapsto L(\sigma^{1/2}L^*(\lambda)^{-1}\sigma^{1/2}) \quad \text{and}$$
$$\kappa_\sigma : \lambda \mapsto L\left(\sigma^{1/2}\frac{1}{e}\exp(-L^*(\lambda))\sigma^{1/2}\right).$$

The respective Jacobians are

$$\nabla h_\sigma\,|_\lambda : \delta \mapsto L(\sigma^{1/2}L^*(\lambda)^{-1}L^*(\delta)L^*(\lambda)^{-1}\sigma^{1/2}) \quad \text{and}$$
$$\nabla\kappa_\sigma\,|_\lambda : \delta \mapsto \frac{1}{e}L(\sigma^{1/2}M_{\exp(-L^*(\lambda))}(-L^*(\delta))\sigma^{1/2}).$$

They are both sign definite as before and, almost verbatim, we can replicate some of the conclusions of Theorems 4 and 5. These are combined into the following statement.

Theorem 6: Let $R \in \mathrm{int}(\mathcal{K})$ and $\sigma \in \mathcal{M}_+$. If $\dim(\mathcal{S}) = 1$, condition (25) holds, and $\lambda_0 \in \mathrm{int}(\mathcal{K}_+^{\mathrm{dual}})$, then the solution to

$$\frac{d}{dt}\lambda_t = (\nabla h_\sigma\,|_{\lambda_t})^{-1}(R - h_\sigma(\lambda_t)) \tag{40}$$

remains in $\mathcal{K}_+^{\mathrm{dual}}$ for $t \geq 0$ and as $t \to \infty$ converges to a unique value $\lambda_{\mathrm{r}} \in \mathcal{K}_+^{\mathrm{dual}}$ such that $R = h_\sigma(\lambda_{\mathrm{r}})$. On the other hand, for any $\lambda_0 \in \mathfrak{R}$, the solution to

$$\frac{d}{dt}\lambda_t = (\nabla\kappa_\sigma\,|_{\lambda_t})^{-1}(R - \kappa_\sigma(\lambda_t)) \tag{41}$$

remains bounded for $t \geq 0$ and as $t \to \infty$ converges to a unique value $\lambda_{\mathrm{e}} \in \mathcal{K}_+^{\mathrm{dual}}$ such that $R = \kappa_\sigma(\lambda_{\mathrm{e}})$. In case $R \notin \mathrm{int}(\mathcal{K})$, the differential equation (41) diverges. In case $R \notin \mathrm{int}(\mathcal{K})$ and $\dim(\mathcal{S}) = 1$, the differential equation (41) diverges as well.

The importance of recasting the relevant parts of Theorems 4 and 5 as above, by incorporating arbitrary σ's in $\mathcal{M}_+$, allows obtaining *any* density function which is consistent with the data R by such a procedure. To see this note that, if ρ is consistent with the data, then working backwards we can select σ accordingly so that ρ equals $\sigma^{1/2}L^*(\lambda)^{-1}\sigma^{1/2}$ or
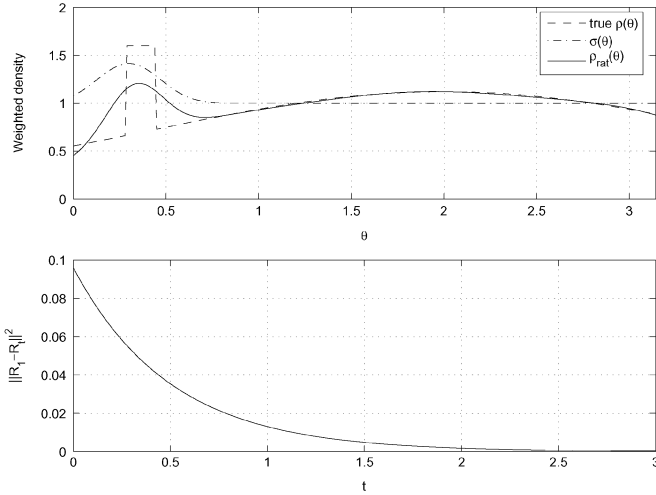
Fig. 3. Computation of spectra for a non-equispaced aarray using a "prior."

$\frac{1}{e}\sigma^{1/2}\exp(-L^*(\lambda))\sigma^{1/2}$ for any λ (in $\mathcal{K}_+^{\text{dual}}$ and $\mathfrak{R}$, respectively). Thus, Theorem 6 gives descriptions of *all* positive densities that are consistent with the data R—simply choose the "correct" σ. Existence of solution per se does not depend on the choice of σ.

A potentially important application is when prior information may dictate a choice of σ. In this case, using Theorem 6 we may obtain an admissible density function which is "closer to our expectations." We amplify this remark below by reworking Example II-A with a suitable weight. We then conclude the section after we revisit and develop Examples II-D and II-E.

A. Nonequispaced Arrays (Example II-A Continued)

Fig. 3 compares the "true" density function ρ_{true} which was used to generate the moments, and a density $\rho_{\text{rat}} = \sigma^{1/2}L^*(\lambda)\sigma^{1/2}$ which is computed according to Theorem 6. The original density ρ_{true} has discontinuous peak at about $\theta \sim 0.35$. Then σ has been selected so as to be ≥ 1 in the neighborhood of $\theta \sim 0.35$—actually centered about 0.25. (The accuracy of the "match" does not seem critical.) The density ρ_{rat} is seen to be a better match as compared with the "unweighted" case of Fig. 2. Subplot (2,1) shows the value of $\|R_1 - R_t\|$ as before, and highlights the fact that, again, ρ_{rat} is consistent with the moments. Since $\lambda_0 G(\theta) \equiv 1$, if we choose $\sigma = \rho_{\text{true}}$ (using 100% hindsight), we obviously obtain a perfect match as explained above.

B. State Statistics and Analytic Interpolation With Degree Constraint (Example II-D Continued)

In Section IV, the map h_σ can be replaced by

$$\lambda \mapsto L(\varphi L^*(\lambda)^{-1}\varphi^*)$$

where $\sigma = \varphi\varphi^*$ is a factorization of σ with φ not necessarily Hermitian, with the obvious modifications in the expression for the corresponding Jacobian. The statement of the theorem holds with no changes. The same applies to κ_σ which can also be cast with respect to an arbitrary factorization of σ—but this will not

concern us here. Instead, we consider the setting of Section II-D where

$$L^*: \lambda \mapsto B^*(I - e^{-j\theta}A^*)^{-1}\lambda(I - e^{j\theta}A)^{-1}B.$$

If we take $\varphi(z) = I + C_o z(I - zA)^{-1}B$ so that φ^{-1} is also analytic in $\mathbb{D}$ (which corresponds to C_o chosen so that $A - BC_o$ is a Hurwitz matrix), then the resulting density function

$$\rho(\theta) = \varphi L^*(\lambda)^{-1}\varphi^*$$
$$= (G_o(e^{j\theta})^*\lambda G_o(e^{j\theta}))^{-1}$$

with $G_o(z) = (I - z(A-BC_o))^{-1}B$. This is a rational spectral density of degree at most twice the dimension of (5), and hence, it gives rise to a positive-real interpolant F as in (9) of McMillan degree at most equal to the dimension of (5).

C. Conforming State Statistics to Underlying Dynamics (Example II-E Continued)

There are several alternative possibilities for the choice of logarithmic entropy distance, e.g., $\mathbb{S}(\hat{R}\|R)$. Yet, as discussed in Section IV, there seems to be no known closed-form expression for the minimizer of $\mathbb{S}(\hat{R}\|R)$. Hence, in analogy with the third alternative given in that section, we seek a value for the state covariance R which minimizes $\mathbb{S}(I\|\hat{R}^{-1/2}R\hat{R}^{-1/2})$ instead.

Conditions (8) locate R in the range of

$$\Gamma: \mathbb{C}^{m\times n} \to \mathbb{C}^{n\times n}: H \mapsto R$$

where $R - ARA^* = BH + H^*B^*$. But if $\{G_i : i = 1, \ldots, p\}$ denotes a set of Hermitian matrices forming a basis for the orthogonal complement of the range of Γ, then R is equally well characterized by the moment-like constraints

$$\langle G_i, R \rangle = 0, \qquad \text{for } i = 1, \ldots, p. \tag{42}$$

Finally, it is both natural and necessary to impose one additional constraint that restricts the size of $\hat{R}^{-1/2}R\hat{R}^{-1/2}$, e.g.,

$$\text{trace}(\hat{R}^{-1}R) = n \tag{43}$$

where n is the size of R. Then the minimizer of $\mathbb{S}(I \| \hat{R}^{-1/2}R\hat{R}^{-1/2})$ subject to (42) and (43) takes the form

$$R = (L^*(\lambda))^{-1} \tag{44}$$

where

$$L^*(\lambda) = \lambda_0\hat{R}^{-1} + \sum_{i=1}^{p}\lambda_i G_i$$

is as usual the adjoint of L that takes

$$R \mapsto [\text{trace}(\hat{R}^{-1}R), \text{trace}(G_1R), \ldots]'.$$

The theory in Section III can be applied to trace a path in the space of Lagrange multipliers from an initial point $\lambda = [1, 0, \ldots, 0]'$ to a final point for which the corresponding R in (44) satisfies (42) and (43).

For a simple numerical example, consider the case where R is sought to be Toeplitz whereas the estimated sample covariance is not, and is given by

$$\hat{R} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Then, the solution described above turns out to be

$$R = \begin{bmatrix} 3.1596 & 1.7394 & 0.9576 \\ 1.7394 & 3.1596 & 1.7394 \\ 0.9576 & 1.7394 & 3.1596 \end{bmatrix}.$$

The methodology works in complete generality for Toeplitz, block-Toeplitz, and structured covariance matrices, and so does an analogous approach based on $\mathbb{S}(R \| \hat{R})$.

V. CONCLUDING REMARKS

We presented an approach for constructing matrix-valued density functions which are consistent with given moments. Section IV describes, in the spirit of the mathematical theory on the moment problem, all positive-definite density functions which are consistent with the data. The nonparametric description given in Section IV (nonparametric since it amounts to an arbitrary choice of a weight-density σ) should prove useful in case we wish to incorporate prior information (e.g., Section IV-A, and cf. [38]).

The basic problem of characterizing admissibilitiy of a matricial moment R has been cast in terms of the positivity of a suitable functional $\mathfrak{C}_R$ in complete analogy with the classical case [51]. However, testing for positivity of such a functional is not a trivial matter. In the classical theory, the "shift" structure of the space of integration kernels ([51], [75], [1], [2]) allows a simple description of all positive elements in their span, via "sums of squares." This is not the case here. Instead, we determine admissibility of R from the convergence of the differential equation given, e.g., in Theorems 4 and 5. Yet, a more direct analog of the Pick operator and a corresponding test that would allow a "certificate of positivity of $\mathfrak{C}_R$," would be highly desirable.

The framework of the present work, when specialized to analytic interpolation as in Sections II-D and IV-B, appears suitable for studying solutions to the most general (tangential and bi-tangential, see [4], [23]) Carathéodory–Nevanlinna–Pick problems with degree constraint. The interpolation directions are specified in general by null directions of an arbitrary inner matrix function $V(z)$ as in (11) and (12), and one is interested in rational solutions with McMillan degree bounded by that of $V(z)$. Earlier work ([28], [6]) dealt only with cases were $V(z)$ is a scalar-inner times the identity. While the main focus of the present work remains the general moment problem, consequences of the theory as in Section IV-B should prove useful in multivariable feedback design with degree constraints (cf. [39]).

APPENDIX I
MATRIX CALCULUS

We assemble a number of basic mathematical formulas. These are expressions for the differential of the matrix exponential and the matrix logarithm that have been used in the physics literature and in quantum information theory.

A. The Matrix Exponential

We begin with the differential of the matrix exponential (see [26], [62]). Following [44, p. 164], integrate both sides of

$$\frac{d}{dt}\left[e^{-tA} e^{t(A+\epsilon B)} \right] = e^{-tA} \epsilon B e^{t(A+\epsilon B)}$$

between 0 and t to obtain

$$e^{-tA} e^{t(A+\epsilon B)} - I = \int_0^t e^{-t_1 A} \epsilon B e^{t_1(A+\epsilon B)} dt_1.$$

Then

$$e^{t(A+\epsilon B)} = e^{tA} + e^{tA} \int_0^t e^{-t_1 A} \epsilon B e^{t_1(A+\epsilon B)} dt_1$$

$$= e^{tA} + e^{tA} \int_0^t e^{-t_1 A} \epsilon B$$

$$\times \left(e^{t_1 A} + e^{t_1 A} \int_0^{t_1} e^{-t_2 A} \epsilon B e^{t_2(A+\epsilon B)} dt_2 \right) dt_1$$

$$= S_0(t) + \epsilon S_1(t) + \epsilon^2 S_2(t) + \dots$$

where

$$S_0(t) = e^{tA}$$

$$S_1(t) = e^{tA} \int_0^t e^{-t_1 A} B e^{t_1 A} dt_1$$

$$S_2(t) = e^{tA} \int_0^t \int_0^{t_1} e^{-t_1 A} B e^{(t_1 - t_2)A} B e^{t_2 A} dt_2 dt_1 \quad (45)$$

and the general term $S_n(t)$ is

$$e^{tA} \int_0^t \int_0^{t_1} \dots \int_0^{t_{n-1}} (e^{-t_1 A} B e^{t_1 A}) \dots$$

$$(e^{-t_n A} B e^{t_n A}) dt_n \dots dt_1.$$

We are only interested in the first two terms of this convergent series.

Evaluating at $t = 1$ and replacing ϵB by Δ, we obtain

$$e^{A+\Delta} - e^A = \int_0^1 e^{(1-\tau)A} \Delta e^{\tau A} d\tau + o(\|\Delta\|).$$

Hence, the differential in the direction Δ (often refered to as Gateaux, or polar, or Fréchet) is given by the linear map

$$\Delta \mapsto \int_0^1 e^{(1-\tau)A} \Delta e^{\tau A} d\tau.$$

This map represents a "scrambled" multiplication of Δ by e^A. To see this assume that A and Δ commute. Then the right-hand side becomes $e^A \Delta$.

The S_2-term in (45) gives the quadratic term in Δ in the expansion of $e^{A+\Delta} - e^A$ as

$$\Delta \mapsto \int_0^1 \left(e^{(1-\tau_1)A} \Delta e^{\tau_1 A} \int_0^1 \left(e^{-A\tau_1 \tau_2} \Delta e^{\tau_1 \tau_2 A} \right) d\tau_2 \right) \tau_1 d\tau_1.$$

In general, for Hermitian matrices C and Δ, and $C \geq 0$, define the "noncommutative" or "scrambled" multiplication of Δ by C via the operator

$$M_C : \Delta \mapsto \int_0^1 C^{(1-\tau)} \Delta C^\tau d\tau. \quad (46)$$

This gives a compact expression for the differential of e^A, summarized below.

Proposition 7: The differential of $\exp(A) := e^A$ is M_{e^A}.

B. The Matrix Logarithm

We now turn to the matrix logarithm. Integrate both sides of

$$\frac{d}{dt}[\log(I+tP) - \log(I+tQ)] = (I+tP)^{-1}P - Q(I+tQ)^{-1}$$

between 0 and 1 to obtain that

$$\log(I+P) - \log(I+Q) = \int_0^1 (I+tQ)^{-1}(P-Q)(I+tP)^{-1}dt$$

assuming that $B := I + P > 0$ and that $A := I + Q >$. Rewrite this expression in terms of A and B and change the integration variable to $\tau = \frac{t-1}{t}$, to obtain

$$\log(B) - \log(A) = \int_0^\infty (A+\tau I)^{-1}(B-A)(B+\tau I)^{-1}d\tau.$$

If $B = A + \Delta$, then

$$(A+\tau I)^{-1} - (A+\Delta+\tau I)^{-1} = (A+\tau I)^{-1}\Delta(A+\Delta+\tau I)^{-1}$$

which, for A, Δ Hermitian, $A > 0$, and $A + \Delta > 0$, leads to

$$\begin{aligned}
&\log(A+\Delta) \\
&= \log(A) + \int_0^\infty (A+\tau I)^{-1}\Delta(A+\tau I)^{-1}d\tau \\
&\quad + \int_0^\infty (A+\tau I)^{-1}\Delta(A+\tau I)^{-1}\Delta(A+\tau I)^{-1}d\tau \\
&\quad + o(\|\Delta\|^2).
\end{aligned} \tag{47}$$

By expanding in terms of eigenvectors of A it can be verified that

$$\int_0^\infty (A+\tau I)^{-1}\Delta(A+\tau I)^{-1}d\tau = M_A^{-1}(\Delta). \tag{48}$$

Indeed, if A is diagonal$\{a_1, \ldots, a_n\}$ then the (i,j)th entry of

$$M_A\left(\int_0^\infty (A+\tau I)^{-1}\Delta(A+\tau I)^{-1}d\tau\right)$$

is simply

$$\int_0^1 a_i^{(1-t)}a_j^t dt \int_0^\infty (a_i+\tau)^{-1}(a_j+\tau)^{-1}d\tau(\Delta)_{ij}$$

where $(\Delta)_{ij}$ is the (i,j)th entry of Δ (in this same basis where A is diagonal). Then

$$\int_0^1 a_i^{(1-t)}a_j^t dt = \frac{a_i - a_j}{\log(a_i) - \log(a_j)}$$

whereas

$$\int_0^\infty (a_i+\tau)^{-1}(a_j+\tau)^{-1}d\tau$$

is the inverse of the same expression. This result is attributed to Lieb (see [74, p. 4]) and summarized below.

Proposition 8: The differential of $\log(A)$ is M_A^{-1}.

REFERENCES

[1] N. I. Akhiezer, *The Classical Moment Problem*. London, U.K.: Hafner Publishing, Oliver and Boyd, 1965.

[2] N. I. Akhiezer and M. G. Krein, *Some Questions in the Theory of Moments*. Providence, RI: Amer. Math. Soc., 1962.

[3] H. Araki, "Relative entropy of states of von Neumann algebras," *Publ. RIMS, Kyoto Univ.*, vol. 11, pp. 809–833, 1976.

[4] 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.

[5] G. A. Bliss, *Lectures on the Calculus of Variations*. Chicago, IL: Univ.Chicago Press, 1946.

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

[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*. Philadelphia, PA: SIAM, June 1994, vol. 15, Studies in Applied Mathematics.

[8] V. B. Braginsky and F. Ya. Khalili, *Quantum Measurement*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[9] J. Burg, "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975.

[10] C. I. Byrnes and A. Lindquist, "A convex optimization approach to generalized moment problems," in *Control and Modeling of Complex Systems: Cybernetics in the 21st Century*. Boston, MA: Birhäuser, 2003, pp. 3–21.

[11] ——, "Interior Point Solutions of Variational Problems and Global Inverse Function Theorems," Roy. Inst. Technol. (KTH), Stockholm, Sweden, Rep. TRITA/MAT01-OS13, 2001.

[12] 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.

[13] ——, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3189–3206, Nov. 2000.

[14] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM Rev.*, vol. 43, no. 4, pp. 645–675, 2001.

[15] ——, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM J. Control and Opt.*, vol. 37, pp. 211–229, 1999.

[16] C. I. Byrnes, T. T. Georgiou, A. Lindquist, and A. Megretski, "Generalized interpolation in H^∞ with a complexity constraint," *Trans. Amer. Math. Soc.*, to be published.

[17] D. Dacunha-Castelle and F. Gamboa, "Maximization de l'entropie sous contraintes non linèaires," *Annales de l'Inst. H. Poincarre*, vol. 4, pp. 567–596, 1990.

[18] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Probab.*, vol. 3, no. 1, pp. 146–158, 1975.

[19] ——, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Probab.*, vol. 19, no. 4, pp. 2032–2066, 1991.

[20] ——, "Sanov property, generalized I-projection and a conditional limit theorem," *Ann. Probab.*, vol. 3, pp. 146–158, 1984.

[21] B. W. Dickinson, "Two dimensional Markov spectrum estimates need not exist," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 1, pp. 120–121, Jan. 1980.

[22] P. Delsarte, Y. Genin, and Y. Kamp, "Half-plane Toeplitz systems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 4, pp. 465–474, Jul. 1980.

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

[24] W. F. Donoghue Jr., *Monotone Matrix Functions and Analytic Continuation*. New York: Springer-Verlag, 1974.

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

[26] R. P. Feynman, "An operator calculus having applications in quantum electrodynamics," *Phys. Rev.*, vol. 84, no. 1, pp. 108–128, 1951.

[27] F. Gamboa and E. Gassiat (Orsay), "Maximum d'entropie et problème des moments: cas multidimensionnel," *Probab. Math. Statist.*, vol. 12, pp. 67–83, 1991.

[28] T. T. Georgiou, "Partial Realization of Covariance Sequences," Ph.D. dissertation, Elec. Eng., Univ. Florida, Gainesville, 1983. Available: [Online] at www.ece.umn.edu/users/georgiou/papers/dissertation.pdf.

[29] ——, "Realization of power spectra from partial covariance sequences," *IEEE Trans. Acous., Speech, Signal Process.*, vol. AASP-35, no. 4, pp. 438–449, Apr. 1987.

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

[31] ——, "A topological approach to Nevanlinna-Pick interpolation," *SIAM J. Math. Anal.*, vol. 18, no. 5, pp. 1248–1260, Sep. 1987.

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

[33] ——, "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.

[34] ——, "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.

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

[36] ——, "Toeplitz covariance matrices and the von Neumann relative entropy," in *Control and Modeling of Complex Systems: Cybernetics in the 21st Century*. Boston, MA: Birkhauser, 2003, pp. 23–30.

[37] ——, *Structured Covariances and Related Approximation Questions (Lecture Notes in Control and Information Sciences)*. New York: Springer-Verlag, 2003, vol. 286, pp. 135–140.

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

[39] ——, "Two alternative views on control design with degree constraint," *IEEE Trans. Autom. Control*, submitted for publication.

[40] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. New York: Halsted Press/Wiley, 1981.

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

[42] S. Haykin, J. H. Justice, N. L. Owsley, J. L. Yen, and A. C. Kak, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[43] J. J. Halliwell, J. Pérez-Mercader, and W. H. Zurek, *Physical Origins of Time Asymetry*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[44] S. P. Heims and E. T. Jaynes, "Theory of gyromagnetic effects and some related magnetic phenomena," *Rev. Mod. Phys.*, vol. 34, no. 2, pp. 143–165, 1962.

[45] D. Jackson, "Formal properties of orthogonal polynomials in two variables," *Duke Math. J.*, vol. 2, pp. 423–434, 1936.

[46] E. T. Jaynes, "On the rationale of maximum entropy methods," *Proc. IEEE*, vol. 70, no. 9, pp. 939–952, Sep. 1982.

[47] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[48] M. Junk, "Maximum entropy for reduced moment problems," *Math. Models and Methods in Appl. Sci.*, vol. 10, no. 7, pp. 1001–1025, 2000.

[49] R. E. Kalman, "Realization of covariance sequences," in *Toeplitz Centennial*, I. C. Gohberg, Ed. Basel, Switzerland: Birkhauser, 1982, pp. 331–342.

[50] C. Kittel, *Thermal Physics*. New York: Wiley, 1969.

[51] M. G. Krein and A. A. Nudel'man, *The Markov Moment Problem and Extremal Problems*. Providence, RI: Amer. Math. Soc., 1977.

[52] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, p. 7986, Mar. 1951.

[53] S. W. Lang and J. H. McClellan, "Multidimensional MEM spectral estimatiion," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 6, pp. 880–887, Dec. 1982.

[54] R. D. Levine and M. Tribus, Eds., *The Maximum Entropy Formalism*. Cambridge, MA: MIT Press, 1979.

[55] G. Le Besnerais, J.-F. Bercher, and G. Demoment, "A new look at entropy for solving linear inverse problems," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1565–1578, Jul. 1999.

[56] A. S. Lewis, "Consistency of moment systems," *Can. J. Math.*, vol. 47, no. 5, pp. 995–1066, 1995.

[57] E. H. Lieb, "Convex trace functions and Wigner-Yanase-Dyson conjecture," *Adv. Math.*, vol. 11, pp. 267–288, 1973.

[58] A. Lesniewski and M. B. Ruskai, "Relative entropy and monotone Riemannian metrics on noncommutative probability spaces," *J. Math. Phys.*, vol. 40, pp. 5702–5724, 1999.

[59] T. L. Marzetta, "Two-dimensional linear prediction: Autocorrelation arrays, minimum-phase predition error filters, and reflection coefficient arrays," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-28, no. 6, pp. 725–733, Dec. 1980.

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

[61] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating poin-process intensities and Toeplitz constrained covariances," *Proc. IEEE*, vol. 75, no. 7, pp. 892–907, Jul. 1987.

[62] I. Najfeld and T. F. Havel, "Derivatives of the matrix exponential and their computation," *Adv. Appl. Math.*, vol. 16, pp. 321–375, 1995.

[63] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

[64] J. von Neumann, "Thermodynamik quantenmechanischer Gesamtheiten," *Nachr. der Gesellschaft der Wiss. Gött.*, pp. 273–291, 1927.

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

[66] M. Ohya and D. Petz, *Relative Entropy and Its Use*. Berlin, Germany: Springer-Verlag, 1993.

[67] D. Petz and Cs. Sudár, "Extending the Fisher metric to density matrices," in *Geometry in Present Days Science*. Singapore: World Scientific, 1999, pp. 21–34.

[68] D. Petz, "Geometry of canonical correlation on the state space of a quantum system," *J. Math. Phys.*, vol. 35, no. 2, pp. 780–795, 1994.

[69] ——, "Quasi-entropies for finite quantum systems," *Rep. Math. Phys.*, vol. 23, no. 1, pp. 57–65, 1986.

[70] F. K. Bose, Ed., "Special Issue on Multidimensional Systems," *Proc. IEEE*, vol. 65, no. 6, Jun. 1977.

[71] M. Rosenblum and J. Rovnyak, *Hardy Classes and Operator Theory*. Oxford, U.K.: Oxford Univ. Press, 1985.

[72] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1987.

[73] ——, "The extension problem for positive-definite functions," *Illinois J. Math.*, vol. 7, pp. 532–539, 1963.

[74] M. B. Ruskai, "Lieb's Simple Proof of Concavity of $(A, B) \mapsto Tr A^p K^\dagger B^{1-p} K$ and Remarks on Related Inequalities," arXiv: quant-ph/0404126v1, Aug. 30, 2004.

[75] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*. New York: Amer. Math. Society, 1943.

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

[77] A. Sudbery, *Quantum Mechanics and the Particles of Nature*. Cambridge, U.K.: Cambridge Univ. Press, 1986.

[78] H. Umegaki, "Conditional expectation in an operator algebra IV. Entropy and Information," *Kodai Math. Sem. Rep.*, vol. 14, pp. 59–85, 1962.

[79] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley-Interscience, 2002, Part IV of Detection, Estimation and Modulation Theory.

[80] J. W. Woods, "Two-dimensional Markov spectral estimation," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 5, pp. 552–559, Sep. 1976.

[81] H. P. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of nonselfadjoint observables," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 740–750, Nov. 1973.