Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Relative Entropy and Moment Problems

Tryphon T. Georgiou

***Abstract*— We consider the von Neumann entropy $\mathbb{I}(\rho) := -\mathrm{trace}(\rho\log\rho)$ and the Kullback-Leibler-Umegaki distance $\mathbb{S}(\rho||\sigma) := \mathrm{trace}(\rho\log\rho - \rho\log\sigma)$ as regularizing functionals in seeking solutions to multi-variable and multi-dimensional moment problems. We show how to obtain extrema for such functionals via a suitable homotopy and how to characterize *all* the solutions to moment problems. The range of possible applications includes the inverse problem of describing power spectra which are consistent with second-order statistics, measurement in classical thermodynamics as well as a quantum mechanics, as well as analytic interpolation encountered in modern robust control (cf. [6], [14], [15], [16], [17]).**

I. Introduction

THE quantum relative entropy (Umegaki [35])

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

where ρ,σ are positive Hermitian matrices (or operators) with trace equal to one, generalizes the Kullback-Leibler relative entropy [23], just as the von Neumann entropy

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

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 [27] 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 [21, Chapter 3])

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

where k runs over all micro-states corresponding to a scalar value $g(k)$. Each micro-state 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., [34, Chapter 5], [19, page 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 [34, page 185]). If the underlying space is infinite dimensional then the measurement process can be modeled via a continuous analogue where the summation is replaced by an integral (e.g., see [4]). 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 $\mathcal{S}\subseteq\mathbb{R}^k$ $(k>1)$ 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, with $d\mu$ a positive Hermitian-valued measure.

The moment problem (1-2) is typified by multivariable and multidimensional sampling of spectra in sensor arrays and polarimetric radar. The echo/signal at different polarizations and/or wavelengths is being sampled at various sensor locations. It is usually the case that these samples are not independent. Attributes of the scattering field (e.g., reflectivity at different wavelengths and polarization) and the relative position of the array elements 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 a $n_{\mathrm{left}}\times m$ transfer matrix

$$G_{\mathrm{left}} = \begin{bmatrix} g_{1,\mathrm{left}} \\ \vdots \\ g_{n_{\mathrm{left}},\mathrm{left}}\end{bmatrix}$$

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

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

represents the vectorial output process. Similarly, if

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

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

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

This work was supported by the NSF and the AFOSR

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455; tryphon@ece.umn.edu

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

$$R = E\{y_{\rm left} y^*_{\rm right}\}$$

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, by using various forms of entropy functionals, we provide a way to answer the following:

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

The present work follows up in the footsteps of a rather extensive literature on inverse problems having roots in the early days of statistical mechanics (e.g., see [25], [11], [24], [26], and also [17] for a more comprehensive list of references). The key idea has been to seek extrema of entropy functionals—existence would guarantee solvability of the moment problem. The idea of using "weighted" entropy functionals to parametrize solutions is more recent and originates in [7]. It was followed up in [6] and in [5], [18] where it was reformulated using the Kullback-Leibler distance between sought solutions and positive "priors." Exploring the connection with the Kullback-Leibler distance, [18], [5], [8], [16], [9] studied *scalar* problems at various levels of generality, while [18] pointed to the relevance of quantum relative entropy for the multi-variable case (see also [12, Chapter V], [2]) Classical moment problems [22] and their matrix-valued counterpart (see e.g., [32]) have been studied when the integration kernels possess a very particular shift-structure similar to that of a Fourier vector and are of limited use in the generality sought herein. In the present work we develop a computational approach for multivariable and multidimensional moment problems analogous to one presented in [16] for scalar distributions, using suitable generalizations of entropy (cf. [18]). Detailed derivations and exposition of the material is provided in [17].

II. Motivating Example: non-uniform sampling

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 Figure 1. For brevity, we have assume a scalar distribution.

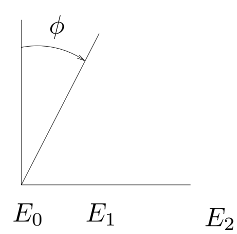

Fig. 1. Non-equispaced sensor array

Assuming that the sensors are sensitive 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 $A(\theta)d\theta$ the amplitude and $\phi(\theta)$ a random phase of the θ-component. Typically, $\phi(\theta)$ for various values of θ are uncorrelated. The term $px_\ell\cos(\theta)$ 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 significance of our selection of distances between sensors, giving rise to the indexing set (3), is to underscore the absence of algebraic dependence between the elements of the *transfer function/array manifold*

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

(thought of as a column vector with $\tau=\cos(\theta)\in[-1,1]$.

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 non-negativity of a Toeplitz 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., [13, page 786], [15]) motivated the present study.

III. Matricial distributions and their moments

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

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

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

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

as a linear space over $\mathbb{R}$. Clearly, both, $\mathfrak{M}$ and $\mathfrak{M}_+$ are convex cones. Since non-negativity 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[1]. 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 space of L does not intersect $\mathfrak{M}$, i.e.,

$$\operatorname{null}(L) \cap \mathfrak{M} = \{0\}. \tag{5}$$

The interior of $\mathcal{K}$ is $\operatorname{int}(\mathcal{K}) = L(\mathfrak{M}_+)$ 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

$$\langle \lambda, R\rangle := \Re e\ (\operatorname{trace}(\lambda^* R)),\ \text{ for } \lambda, R \in \mathfrak{R}. \tag{6}$$

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

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

The interior of the dual cone

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

corresponds to $\mathfrak{M}_+$ since $\mathcal{K}_+^{\text{dual}} = \{\lambda\ :\ L^*(\lambda) \in \mathfrak{M}_+\}$. Finally, (5) can be seen to be equivalent to $\mathcal{K}_+^{\text{dual}} \neq \emptyset$.

[1] 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 [22]).

*1) **Minimizers of** $\mathbb{S}(I\|\rho)$**:*** We are interested in minimizers of (the negative entropy) $\mathbb{S}(I\|\rho) = -\operatorname{trace}(\log(\rho))$ on $\mathfrak{M}_+$ subject to $R = L(\rho)$. 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 $-\operatorname{trace}(\log(\rho))$).

The Lagrangian of the problem is

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

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

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

In the above derivation, the "trace" is what allows replacing the "non-commutative division operator" M_ρ^{-1} (cf. (32)) with multiplication by ρ^{-1}. The stationarity condition $d\mathcal{L}(\lambda, \rho\,;\, \delta) \equiv 0$ then gives

$$\rho \;=\; (L^*(\lambda))^{-1}. \tag{7}$$

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 $R \in \operatorname{int}(\mathcal{K})$ then this condition is also sufficient as claimed below.

Theorem 1: Assume that $R \in \operatorname{int}(\mathcal{K})$. 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.

*2) **Minimizers of** $\mathbb{S}(\rho\|I)$**:*** We now focus on minimizers of $\mathbb{S}(\rho\|I) = \operatorname{trace}(\rho\log(\rho))$ in $\mathfrak{M}_+$, subject to $R = L(\rho)$. The Lagrangian this time is

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

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

$$\begin{aligned} d\mathcal{L}(\lambda, \rho\,;\, \delta) &:= \operatorname{trace}(\delta\log(\rho) + \rho M_\rho^{-1}(\delta)) + \langle \lambda, L(\delta)\rangle\\ &= \operatorname{trace}(\delta\log(\rho) + \delta) + \langle L^*(\lambda), \delta\rangle. \end{aligned}$$

The last step follows from

$$\begin{aligned} \operatorname{trace}(\rho M_\rho^{-1}(\delta)) &= \operatorname{trace}(\rho\int_0^\infty (\rho + t)^{-1}\delta(\rho + t)^{-1}dt)\\ &= \operatorname{trace}(\int_0^\infty (\rho + t)^{-1}\rho(\rho + t)^{-1}dt\,\delta) = \operatorname{trace}(\delta). \end{aligned}$$

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

$$\rho = \exp\left(-I - L^*(\lambda)\right) = \frac{1}{e}\exp\left(-L^*(\lambda)\right) \tag{8}$$

with $L^*(\lambda) \in \mathbb{M}$ (and not necessarily in $\mathfrak{M}$ as before). It turns out that if $R \in \operatorname{int}(\mathcal{K})$ a minimizer can always be found. It should be noted that (5) is no longer a necessary condition.

Theorem 2: If $R \in \operatorname{int}(\mathcal{K})$, then the entropy functional $\mathbb{S}(\rho\|I)$ has a minimum in $\mathfrak{M}_{R,+}$ which is unique and of the form (8).

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 (1-2), 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 multi-dimensional 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

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

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

$$L \;:\; \tilde{\mathfrak{M}} \to \mathfrak{R} \;:\; \rho \mapsto R = \int_{\mathcal{S}} G_{\rm left}(\theta)\rho(\theta)G_{\rm right}(\theta)d\theta \quad (9)$$

maps $\tilde{\mathfrak{M}}$ into a subspace of $\mathbb{C}^{\rm left \times right}$ 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 (6). For simplicity of the exposition, we assume that the integration kernels $G_{\rm left}, G_{\rm right}$ are *continuously differentiable* on $\mathcal{S}$. The *closure* of the range of $\tilde{\mathfrak{M}}$ is denoted by $\mathcal{K} = \overline{L(\tilde{\mathfrak{M}})}$, while $\operatorname{int}(\mathcal{K}) = L(\tilde{\mathfrak{M}}_+)$. The adjoint transformation is now

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

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

$$\begin{aligned}\mathcal{K}^{\rm dual} &= \{\lambda \in \mathfrak{R} \;:\; L^*(\lambda) \in \tilde{\mathfrak{M}}\}, \text{ and}\\ \mathcal{K}^{\rm dual}_+ &= \{\lambda \in \mathfrak{R} \;:\; L^*(\lambda) \in \tilde{\mathfrak{M}}_+\}\end{aligned}$$

remain valid (except for the obvious change where $\tilde{\mathfrak{M}}$ replaces our earlier $\mathfrak{M}$). The analog of (5) will be needed (in Theorem 4) which, can also be expressed as

$$K^{\rm dual}_+ \neq \emptyset. \quad (10)$$

Finally we define as before

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

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

For future reference we bring in a characterization of elements $R \in \mathcal{K}$ analogous to the scalar real case given in [22, page 14]. Given $R \in \mathfrak{R}$, define the real-valued functional

$$\mathfrak{C}_R \;:\; \mathfrak{R} \to \mathbb{R} \;:\; \lambda \mapsto \langle \lambda, R\rangle \quad (11)$$

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 $\lambda \in \mathcal{K}^{\rm dual}_+$ of unit norm is positive (resp. nonnegative).

Proposition 3: The following hold:

$$\begin{aligned} R \in \mathcal{K} &\Leftrightarrow \mathfrak{C}_R \geq 0\\ R \in \operatorname{int}(\mathcal{K}) &\Leftrightarrow \mathfrak{C}_R > 0.\end{aligned}$$

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

$$\tilde{\mathbb{S}}(\rho\|\sigma) := \int_{\mathcal{S}} \operatorname{trace}(\rho \log \rho - \rho \log \sigma) d\theta. \quad (12)$$

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, $\tilde{\mathbb{S}}(I\|\rho)$ and then $\tilde{\mathbb{S}}(\rho\|I)$. 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 $\tilde{\mathbb{S}}(I\|\rho)$.

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

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

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

$$\begin{aligned} d\mathcal{L}(\lambda, \rho\,;\,\delta) &:= \operatorname{trace} \int_{\mathcal{S}} (-M^{-1}_{\rho(\theta)} + L^*(\lambda))\delta(\theta)d\theta\\ &= \operatorname{trace}\int_{\mathcal{S}} (-\rho(\theta)^{-1} + L^*(\lambda))\delta(\theta)d\theta,\end{aligned}$$

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

$$\rho = L^*(\lambda)^{-1}. \quad (13)$$

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

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

where we seek a solution to the moment constraints (2). It turns out that if a solution exists then, a particular one exists in $\tilde{\mathfrak{M}}_{\rm rat}$ and that it can be obtained by computing the fixed point of an exponentially converging matrix differential equation as stated below.

Theorem 4: If $\dim(\mathcal{S}) = 1$, condition (10) holds, and $R \in \operatorname{int}(\mathcal{K})$, then $\tilde{\mathbb{S}}(I\|\rho)$ has a minimum in $\tilde{\mathfrak{M}}_{R,+}$ which is unique and belongs to $\tilde{\mathfrak{M}}_{\rm rat}$. Furthermore, for any $\lambda_0 \in \mathcal{K}^{\rm dual}_+$, the solution λ_t of the matrix differential equation

$$\frac{d}{dt}\lambda_t = \left(\nabla h|_{\lambda_t}\right)^{-1}(R - L(L^*(\lambda_t)^{-1})), \quad (14)$$

where

$$\nabla h|_{\lambda_t} : \mathfrak{R} \to \mathfrak{R} \;:\; \delta \mapsto L(L^*(\lambda_t)^{-1}L^*(\delta)L^*(\lambda_t)^{-1}), \quad (15)$$

belongs to $\mathcal{K}^{\rm dual}_+$ for all $t \in [0,\infty)$, it converges to a point $\hat{\lambda} \in \mathcal{K}^{\rm dual}_+$ as $t \to \infty$ corresponding to this unique minimizer $\rho = L^*(\hat{\lambda})^{-1}$ for $\tilde{\mathbb{S}}(I\|\rho)$ satisfying $R = L(\rho)$. The differential equation (14) is exponentially convergent as the square distance $V(\lambda_t) = \|R - L(L^*(\lambda_t)^{-1})\|^2$ satisfies $dV(\lambda_t)/dt = -2V(\lambda_t)$. Conversely, if $R \notin \operatorname{int}(\mathcal{K})$ and the dimension of $\mathcal{S}$ is one, then the differential equation (14) diverges.

Equation (14) is equivalent to

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

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 (14) appears preferable for numerical reasons.

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

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

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

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

The stationarity condition leads to the expression

$$\rho = \frac{1}{e}\exp(-L^*(\lambda)), \tag{17}$$

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

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

of potential minimizers for $\tilde{\mathbb{S}}(\rho\,\|I)$, where we seek a solution to (2). The development runs in parallel to the case where $\rho \in \tilde{\mathfrak{M}}_{\mathrm{rat}}$ with one important difference. The "Lagrange multipliers" λ no longer need to be restricted to $\mathcal{K}^{\mathrm{dual}}$ and existence of solutions when $R \in \mathcal{K}$ can be guaranteed even when $\dim(\mathcal{S}) > 1$. Moreover, (10) is no longer necessary and existence of solution to the moment problem in $\tilde{\mathfrak{M}}_{\exp}$ is impervious to the dimension of the dual cone $\mathcal{K}_+^{\mathrm{dual}}$.

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

$$\frac{d}{dt}\lambda_t = \left(\nabla k|_{\lambda_t}\right)^{-1}(R - L(\lambda_t)), \tag{18}$$

where

$$\nabla k|_{\lambda_t} : \mathfrak{R} \to \mathfrak{R} \;:\; \delta \mapsto -\frac{1}{e}L(M_{\exp(-L^*(\lambda_t))}(L^*(\delta)), \tag{19}$$

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 $\tilde{\mathbb{S}}(I\|\rho)$ 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 $R \notin \operatorname{int}(\mathcal{K})$ then the differential equation (18) diverges.

C. Non-equispaced arrays (cont.)

We continue with Example II. We begin with a "true" density ρ_{true} shown in Figure 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 (14) and (18) taking $\lambda_0 = [1\ \ 0\ \ 0\ \ 0]$, and display in Figure 2 the resulting $\rho_{\exp}(\lambda_\infty,\theta)$ and $\rho_{\mathrm{rat}}(\lambda_\infty,\theta)$, for comparison. Both are constructed using the fixed point of the corresponding differential equations.

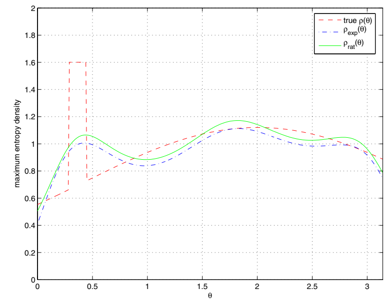

Fig. 2.

IV. The complete set of positive solutions

Reference [18] suggested that all positive solutions to the moment problem may be obtained as minimizers of a suitable entropy functional, e.g., as being

$$\operatorname{argmin}\{\mathbb{S}(\sigma\,\|\rho) \;:\; R = L(\rho)\} \tag{20}$$

with σ thought of as a parameter. This was carried out successfully in [18] and [16] 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 non-constant case is straightforward and a positive result is given for the general case.

Considering the Lagrangian and the stationarity conditions for (20) we arive at

$$\begin{aligned} d\mathcal{L}(\lambda,\rho\,;\,\delta) &= \operatorname{trace}(-\delta M_\rho^{-1}(\sigma)) + \langle L^*(\lambda),\delta\rangle \\ &= \operatorname{trace}\left(-\delta M_\rho^{-1}(\sigma) + \delta L^*(\lambda)\right), \end{aligned}$$

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 (20) is unkown.

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

$$\begin{aligned} h_\sigma &\;:\; \lambda \mapsto L(\sigma^{1/2}L^*(\lambda)^{-1}\sigma^{1/2}),\ \text{and} \\ \kappa_\sigma &\;:\; \lambda \mapsto L(\sigma^{1/2}\frac{1}{e}\exp(-L^*(\lambda))\sigma^{1/2}). \end{aligned}$$

The respective Jacobians are

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

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

Theorem 6: Let $R \in \mathrm{int}(\mathcal{K})$ and $\sigma \in \tilde{\mathfrak{M}}_+$. If $\dim(\mathcal{S}) = 1$, condition (10) holds, and $\lambda_0 \in \mathrm{int}(\mathcal{K}_+^*)$, then the solution to

$$\frac{d}{dt}\lambda_t = \left(\nabla h_\sigma|_{\lambda_t}\right)^{-1}(R - h_\sigma(\lambda_t)) \tag{21}$$

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

$$\frac{d}{dt}\lambda_t = \left(\nabla \kappa_\sigma|_{\lambda_t}\right)^{-1}(R - \kappa_\sigma(\lambda_t)) \tag{22}$$

remains bounded for $t \geq 0$ and as $t \to \infty$ converges to a unique value $\lambda_{\rm e} \in \mathcal{K}_+^{\rm dual}$ such that $R = \kappa_\sigma(\lambda_{\rm e})$. In case $R \notin \mathrm{int}(\mathcal{K})$, then (22) diverges. In case $R \notin \mathrm{int}(\mathcal{K})$ and $\dim(\mathcal{S}) = 1$, then (21) diverges as well.

The importance of recasting Theorems 4 and 5 as above, by incorporating arbitrary σ's in $\tilde{\mathfrak{M}}_+$, allows obtaining *any* density function which is consistent with the data R by such a procedure. To see this note that, if ρ 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 $\frac{1}{e}\sigma^{1/2}\exp(-L^*(\lambda))\sigma^{1/2}$ for any λ (in $\mathcal{K}_+^{\rm dual}$ 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.

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" (see [17]). Another interesting usage of the above is, in the more structured setting of analytic interpolation problems, to characterize a class of solutions with a degree constraint, as we briefly explain next.

A. Analytic interpolation with degree constraint

Consider the linear discrete-time state equations

$$x_k \;=\; Ax_{k-1} + Bu_k, \text{ for } k \in \mathbb{Z}, \tag{23}$$

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 non-negative 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

$$R = \int_{-\pi}^{\pi}\left(G(e^{j\theta})\frac{d\mu(\theta)}{2\pi}G(e^{j\theta})^*\right) \tag{24}$$

where $G(z) := (I - zA)^{-1}B$. is the transfer function of system (23), and is characterized by the following two equivalent conditions (see [15])

$$\mathrm{rank}\begin{bmatrix} R - ARA^* & B \\ B^* & 0 \end{bmatrix} = 2m \tag{25}$$

and,

$$R - ARA^* = BH + H^*B^* \text{ for some } H \in \mathbb{C}^{m\times n}. \tag{26}$$

Power spectral measures consistent with (24) 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

$$F(z) = \int_{-\pi}^{\pi}\left(\frac{1 + ze^{j\theta}}{1 - ze^{j\theta}}\right)\frac{d\mu(\theta)}{2\pi} + jc, \tag{27}$$

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.,

$$d\mu(\theta) \sim \lim_{r\nearrow 1}\Re(F(re^{j\theta})). \tag{28}$$

The class of nonnegative real matrix valued functions F giving rise to admissible power spectral measures are also characterized by the interpolation condition ([15])

$$F(z) = H(I - zA)^{-1}B + Q(z)V(z) \tag{29}$$

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

$$V(z) := D + zC(I - zA)^{-1}B \tag{30}$$

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 equation (29) specify an analytic interpolation problem of the Nehari type. Positive-real solutions to (29) can be given via (27) and solutions to the moment problem (24). The characterization of solutions to (24) given in Theorem 6 allows a non-classical characterization of solutions to (29) and in particular a characterization of solutions of McMillan degree less than or equal to the dimension of (23). In fact, the map h_σ in Section IV, can be rewritten as

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

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 since we focus on rational solutions.

In the current setting,

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

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

$$\begin{aligned}\rho(\theta) &= \varphi L^*(\lambda)^{-1}\varphi^* \\ &= (G_o(e^{j\theta})^*\lambda G_o(e^{j\theta}))^{-1},\end{aligned}$$

with $G_o(z) = (I - z(A - BC))^{-1}B$. This is a rational spectral density of degree at most twice the dimension of (23), and hence, it gives rise to a positive-real interpolant F of McMillan degree at most equal to the dimension of (23).

V. Concluding remarks

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 [30], [29]. In particular Araki's theory [1], [28] helped characterize a family of "quasi-entropies," contractive under stochastic maps. References [31], [28] 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_W in Section IV, are metrics induced by a quasi-entropy in the language of Petz [31]. Finally, a counterpart for discrete distributions relates to the theory of analytic centers in semi-definite programming, e.g., see [3]. 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.

The author would like to thank Pablo Parillo for technical input, and Laurent Baratchart and Anders Lindquist for inspiring discussions.

VI. Appendix: Differential of $\exp(\cdot)$ and $\log(\cdot)$

We assemble certain formulae for the differentials of the matrix exponential and logarithm. These seem to be largely unknown in the controls literature. Following [20, page 164] (see also [10]),

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

and the differential of the exponential in the direction Δ (Fréchet) is given by

$$M_C \ : \ \Delta \mapsto \int_0^1 C^{(1-\tau)}\Delta C^{\tau} d\tau \tag{31}$$

and $C = e^A$. The map M_C represents a "non-commutative multiplication" of C with Δ. Similarly, up to $o(\|\Delta\|)$,

$$\begin{aligned}\log(A+\Delta) - \log(A) \ &\sim \ \int_0^\infty (A+\tau I)^{-1}\Delta(A+\tau I)^{-1} d\tau \\ &= \ M_A^{-1}(\Delta).\end{aligned} \tag{32}$$

Either expression represents the differential of $\log(A)$ (see [33, page 4]).

References

[1] H. Araki, "Relative entropy of states of von Neumann algebras," *Publ. RIMS, Kyoto Univ.*, **11**: 809-833, 1976.

[2] A. Blomqvist, A. Lindquist, and R. Nagamune, "Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach," *IEEE Trans. on Automatic Control,* **48(12)**: 2172-2190, December 2003.

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

[4] V.B. Braginsky and F.Ya. Khalili, Quantum measurement, Cambridge Univ. Press, 1992.

[5] C. I. Byrnes and A. Lindquist, "A convex optimization approach to generalized moment problems," in **Control and Modeling of Complex Systems**, pages 3-21, Birhäuser, 2003.

[6] 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. on Automatic Control* **45(6):** 822-839, June 2001.

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

[8] C.I. Byrnes, T.T. Georgiou, A. Lindquist, and A. Megretski, "Generalized interpolation in H^∞ with a complexity constraint," *Trans. of the American Math. Society*, to appear in 2005.

[9] P. Enqvist,"A homotopy approach to rational covariance extension with degree constraint," *Intern. J. Appl. Math. and Comp. Science* **11(5)** (2001), 1173–1201.

[10] R.P. Feynman, "An operator calculus having applications in quantum electrodynamics," Physical Review, **84(1)**: 108-128, 1951.

[11] F. Gamboa et E. Gassiat (Orsay), "Maximum d'entropie et problème des moments: cas multidimensionnel," *Probability and Mathematical Statistics,* **12**, Fasc. 1, 67-83, 1991.

[12] T.T. Georgiou, **Partial Realization of Covariance Sequences,** Ph.D. Dissertation, University of Florida, 1983; www.ece.umn.edu/users/georgiou/papers/dissertation.pdf

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

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

[15] T.T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. on Automatic Control*, **47(7):** 1056-1066, July 2002.

[16] T.T. Georgiou, "Solution of the general moment problem via a one-parameter imbedding," *IEEE Trans. on Automatic Control*, to appear, `www.ece.umn.edu/users/georgiou/papers/AC_04_140_final.pdf`

[17] T.T. Georgiou, "Relative Entropy and the multi-variable multi-dimensional Moment Problem," *IEEE Trans. on Information Theory*, accepted for publication; preprint arXiv:math.OC/0506124v1.

[18] T.T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. on Information Theory*, **49(11)**, November 2003.

[19] J.J. Halliwell, J. Pérez-Mercader, and W.H. Zurek, Physical origins of time asymetry, Cambridge Univ. Press, 1994.

[20] S.P. Heims and E.T. Jaynes, "Theory of gyromagnetic effects and some related magnetic phenomena," Reviews of Modern Physics, **34(2)**: 143-165, 1962.

[21] C. Kittel, Thermal Physics, Wiley, 1969.

[22] M.G. Krein and A.A. Nudel'man, **The Markov Moment Problem and Extremal Problems,** American Mathematical Society, Providence, RI, 417 pages, 1977.

[23] S. Kullback and R.A. Leibler, "On information and sufficiency," Ann. Math. Statistics, 1951.

[24] S.W. Lang and J.H. McClellan, "Multidimensional MEM spectral estimatiion," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-30(6)**: 880-887, 1982.

[25] G. Le Besnerais, J-F. Bercher, and G. Demoment, "A new look at entropy for solving linear inverse problems," *IEEE Trans. on Inform. Theory*, **45(5)**: 1565-1578, 1999.

[26] A.S. Lewis, "Consistency of moment systems," *Can. J. Math.*, **47(5)**: 995-1066, 1995.

[27] E.H. Lieb, "Convex trace functions and Wigner-Yanase-Dyson conjecture," Advances in Mathematics, 11: 267-288, 1973.

[28] A. Lesniewski and M.B. Ruskai, "Relative entropy and monotone Riemannian metrics on noncommutative probability spaces," *J. Math. Physics*, **40**: 5702-5724, 1999.

[29] M. Ohya and D. Petz, **Relative Engropy and its Use**, Springer-Verlag, 1993.

[30] D. Petz and Cs. Sudár, "Extending the Fisher metric to density matrices," in *Geometry in Present Days Science*, eds. O.E. Barndorff-Nielsen and E.B. Vendel Jensen, pages 21-34, World Scientific, 1999.

[31] D. Petz, "Geometry of canonical correlation on the state space of a quantum system," *J. Math. Phys.*, **35(2)**: 780-795, 1994.

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

[33] M.B. Ruskai, "Lieb's simple proof of concavity of $(A,B) \mapsto Tr\ A^pK^\dagger B^{1-p}K$ and remarks on related inequalities," arXiv: quant-ph/0404126v1, August 30, 2004.

[34] A. Sudbery, Quantum mechanics and the particles of nature, Cambridge Univ. Press, 1986.

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