Solution of the general moment problem via a one-parameter imbedding

Tryphon T. Georgiou, *IEEE Fellow*

Abstract— **This paper presents a computational theory for the general scalar moment problem. The formalism is sufficiently general to encompass problems in sensor arrays with arbitrary geometry and dynamics, and in nonuniform multidimensional sampling.**

Given a finite set of moments, the theory provides a test for the existence of a positive measure which is consistent with such data. At the same time the theory also provides a characterization of all such consistent positive measures. It should be noted that classical results (e.g., in the theory of the trigonometric moment problem, Hamburger, Stieljes, Nevanlinna-Pick interpolation, etc.) are not applicable to the general setting sought herein where there is no natural shift operator in the space spanned by the integration kernels.

The centerpiece of the theory is a differential equation which depends on the given finite set of moments and on an arbitrary positive function Ψ—which plays the role of a "free parameter". The differential equation has an exponentially attractive point of equilibrium if and only if there exists a consistent positive measure. For each Ψ, the fixed point determines a corresponding measure. Suitable choice of Ψ allows recovering any measure which is consistent with the data. The fixed point of the differential equation corresponds to an extremum of an entropy-like functional, and the differential equation is constructed via an appropriate homotopy that follows changes in the Lagrange multipliers from a convenient starting value to a value for the multipliers that corresponds to the given moments.

Keywords— **Sensor arrays, antenna arrays, multidimensional spectral analysis, multidimensional moment problem, relative entropy, Kullback-Leibler distance, homotopy.**

I. Introduction

In 1894 Stieljes published his classical memoir [56]: "Recherches sur les fractions continue," where he posed and solved the following problem: find a bounded nondecreasing function $\mu(\theta)$ on $[0,\infty)$ such that its "moments" $\int_0^\infty \theta^n d\mu(\theta)$ have specified values for $n = 0, 1, 2, \ldots$. The term "moments" was borrowed from Mechanics. Subsequent formulations dealt with the support and properties of the distribution function μ, the class of integration kernels (e.g. $\{1, \theta, \theta^2, \ldots\}$ above), and the cardinality of this set. At the present time, more than a century later, a vast amount of literature has grown around this type of a problem.

The classical theory was developed for function spaces where a convenient representation of positive elements is available as a sum of squares. Such a property is convenient when testing for the solvability of the moment problem—the existence of a solution relates to the sign definiteness of a suitable quadratic form (Toeplitz, Pick, etc.). In the mid 1930's, M.G. Krein discovered a deep connection with the theory of convex bodies and much of the classical theory was extended to Tchebysev systems (see e.g., the classical monographs by the Russian school [1], [2], [39], [53], and also [37]). Deep connections with the analytic interpolation theory of the early part of the twentieth century (Carathéodory, Schur, Nevanlinna, Pick and others, see [33]) were explored and a beautiful operator theory has emerged (see e.g., [52], [6], [3], [4], [19] and [5]).

The multi-dimensional moment problem where the support of $d\mu$ is of dimension larger than one, has led a somewhat separate existence. The dichotomy between the algebras of polynomials in one and, in many variables, is a major difference and as a consequence the corresponding operator theory is far deeper (e.g., [49], [16]). In the absence of a manageable algebraic structure [50], [18], [45], a computational approach was sought based on the use of entropy functionals [58], [41], [47], [17], [22], [44], [43], [36], [21]. Entropy functionals are natural barriers on convex cones, and hence, the idea is to seek a positive measure which is an extremum of such a functional. The present paper builds on this general idea, yet it follows a distinct path, which has led to further multivariable generalizations [30], [29].

Two logarithmic entropy functionals have received particular attention, $-\int \log m(\theta) d\theta$ and $\int m(\theta) \log m(\theta) d\theta$. Minimization of the first, subject to moment constraints, leads to a "rational" model. Minimization of the second leads to an "exponential" one. There is usually a preference for the first one when m is the spectral density function of a random process, and for the second when m is a probability density function (pdf). This is due to a natural interpretention of such integrals in the respective contexts ([40], [42]) with $\int \log(m(\theta)) d\theta$ being entropy rate while $\int m(\theta) \log m(\theta) d\theta$ simply the entropy of a random variable. Yet, at times a direct interpretation may be unnecessary ([47], [32]) and the functional simply thought of as a computational tool. To this end, alternative choices with similar properties are $-\int \Psi(\theta) \log(m(\theta)) d\theta$ and $\int \Psi(\theta) m(\theta) \log m(\theta) d\theta$, for added flexibility in selecting a "weight" Ψ".

The existence of extrema has been investigated in [22], [44], [7], [8], [36] with tools from probability and large deviations theory, convex optimization, and duality theory, in great generality. Invariably, computations are carried out by seeking the extremum of a dual unconstrained functional via a (carefully stepped up) Newton method. In contrast, our starting point is an approach suggested in [23, page 76], to use a homotopy in the space of moments. This allows us to follow a corresponding path of extrema in the space of Lagrange multipliers. It provides an indepen-

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455; tryphon@ece.umn.edu. Research supported by NSF and AFOSR

dent treatment of the existence of extrema that determines the solvability of the moment problem. The Lagrange multipliers along the path obey a differential equation which converges with an a-priori guaranteed rate, provided the moment problem is solvable. Otherwise the differential equation diverges.

The same tools can be used to characterize the complete solution set of positive measures. The functional form of extrema for a relative entropy functional is taken instead. This depends on Lagrange multipliers and on a "parameter" Ψ (a positive function with the same support as the $d\mu$'s). The differential equation is constructed accordingly. When it converges, then it does so independently of our choice of Ψ. Selecting different Ψ's allows us to recover *all* (absolutely continuous) positive measures that are consistent with the given moments.

An important observation is that a diffeomorphism between moments and Lagrange multipliers (under certain conditions, see [36, Corollary 6.1], [13], and Section V below) can be the basis of our theory. In the present work on the scalar moment problem such a diffeomorphism is given to us by the functional form of the entropy extrema. As shown in a follow-up paper [30], the theory carries through in a similar manner in the multivariable setting as well.

Works that have influenced our current development include [12] who studied an entropy functional as a computational tool for the "rational covariance extension problem" ([23], [24]), [14] who study in a similar vain existence of rational families for a general moment problems, [15] for interpolation problems, and [31], [14] who suggest a relative entropy functional for seeking minimizers in rational form. Early on, [41] also studied rational models for power spectra via an entropy functional in the context of antenna arrays. We will develop both rational and exponential models, and as we will see below, it is the latter that applies in complete generality in higher dimensions.

We begin with three motivating examples in Section II. These are revisited in Section VI after the exposition of the main technical results in Section III. Section III is followed by an analytical example in Section IV that provides insight to the derivation of the main results in Section V.

II. Motivating examples

We begin with three basic examples that underscore the relevance and applicability of the theory that follows. After we explain the key elements of the theory (in Theorems 2 and 3 and in Section III), we will revisit these examples and work them out with specific numerical values.

A. Input power spectrum from output measurements

Consider that measurements of a certain continuous-time, zero-mean, stationary stochastic process $\{u(t) : t \in \mathbb{R}\}$ are obtained at the output of several (low-pass) sensors with time-constant τ_k ($k \in \{1, 2, \ldots\}$). Let $u_k(t)$ denote the corresponding outputs and assume that only variances $r_k = E\{u_k(t)^2\}$ are available as well as $r_0 = E\{u(t)^2\}$.

If $R(\tau)$ denotes the (continuous) autocorrelation function of $u(t)$ and $d\mu$ denotes the non-negative, finite spectral measure of the process, the two are related via

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega} d\mu(\omega).$$

If $H_k(i\omega) = \frac{1}{i\omega/\tau_k+1}$ denotes the transfer function of the typical sensor, then the r_k's ($k \in \{0, 1, 2, \ldots\}$) represent moments of the spectral measure $d\mu(\omega)$ for kernel functions $g_k(\omega) := |H_k(i\omega)|^2$. Indeed, the covariances at the output of the sensors satisfy

$$r_k = \int_{-\infty}^{\infty} g_k(\omega) d\mu(\omega) \text{ for } k = 0, 1, \ldots, n,$$

with

$$g_k(\omega) = \frac{1}{\omega^2/\tau_k^2 + 1},\ k = 1, 2, 3, \ldots \text{ and } g_0 = 1.$$

It is not hard to imagine a situation where statistics can be collected for an assortment of variables for which we possess information on their dynamical link with an unknown "input" stochastic process. Such statistics are precisely moment constraints on the spectral measure of the input. The natural question then arises as to *what we can infer about the spectral measure of the input based on a finite set of covariance statistics*, such as the set $\{r_0, r_1, r_2, r_3, \ldots\}$.

The discrete-time counterpart of this question, has been raised and addressed in [10], [11] (see also [26] and [27], [28]). In that, there is a natural "shift" operator that relates the integration kernels. Such a relationship between the kernels $g_k(\omega)$ is absent in our present "continuous-time" setting. As a consequence, the results in [10], [11], [26] and [27] are not directly applicable.

B. Non-uniformly spaced sensor arrays

In a variety of applications, typified by radar, sonar, and various ultrasound imaging apparatus', the analysis of impinging waves on an array of sensors reveals the makeup of the scattering medium. Propagation delays in the medium translate into sensor-dependent phase-shifts. The goal is then to unravel the effect of propagation delays and assess the distribution of incoming energy from different directions.

The special case of a uniform linear array falls within the setting of standard sampling theory (e.g., see [35, Section 2]). The signal power, as a function of the angular direction of origin, can be shown to be a moment generating function of the sample correlations between the various sensors. Fourier techniques are then suitable and, hence, extensively used. However, a challenging situation arises when the array is not linear and/or not regularly spaced.

Situations of sensors with complicated geometry abound. For instance, consider an array of sonar buoys launched at some point in time and drifting along while their position may be continuously monitored via GPS. In other cases, while we may be able to control the position of the sensors, these are expensive and limited in number. Then it may happen that a non-uniform positioning offers advantages (as is the case for the "Y" shape of the VLA radio telescope

in New Mexico, see [35, page 86]). In situations where the geometry is complicated, Fourier techniques are of little use. Instead, model based methods, beamforming, and the sensitivity of the array as a function of direction, help to interpret the recorded signal and its statistics. For a state-of-the art account of sensor array processing we refer to [57] (see also [54], [34], [35]).

For our purposes, it suffices to consider a two-dimensional planar monochromatic wave impinging upon an array of three sensors. The three elements of this array (sensors) are indicated as E_1, E_2 and E_3 in Figure 1. They are placed linearly but in non-commensurable distances from one another. In particular we may take those distances to be 1 and π wavelengths, as exemplified in Figure 1. The choice of "π" only serves to underscore that the distances are non-commensurable. The waves are assumed to have support in the sector $(0, \frac{\pi}{2})$, highlighted in the upper part of the figure. A simple model for the waves impinging at the k^{th} element is

$$u_k(t) = \int_0^{\pi/2} A(\phi) \sin(\omega t - \kappa x_k + \psi(\phi)) d\phi$$

with $A(\phi)$ the amplitude in the direction ϕ, $\psi(\phi)$ a random phase uniformly distributed over ϕ and independent for different values of ϕ, κ the wavenumber, and x_k the distance of the k^{th} element from the first one (chosen to be the element E_1, hence $x_1 = 0$, $x_2 = 1$ and $x_3 = 1 + \pi$).

By correlating the readings at the three locations, two at a time, we arrive at the correlations

$$\begin{aligned} r_\ell &= \int_0^{\pi/2} \frac{1}{2} \cos(\ell\kappa \sin(\phi)) A(\phi)^2 d\phi \\ &= \int_0^1 \cos(\ell\theta) d\mu(\theta) \end{aligned} \tag{1}$$

for $\ell \in \{0, 1, \pi, \pi+1\}$, where in the last equation we re-scale $\theta := \sin(\phi)$, we set $\kappa = 1$, and we absorb the nonlinear scaling and the amplitude into $d\mu(\theta)$. The index ℓ is chosen to represent the distance between the respective sensors whose outputs are being correlated. It is quite natural to

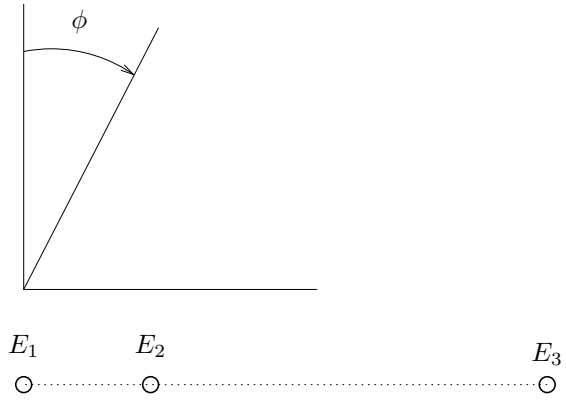

Fig. 1. Geometry of sensor array

ask the following basic questions:

(a) *given real values $r_0, r_1, r_\pi, r_{\pi+1}$ how can we tell whether they are admissible correlation values for the setting that we have just described?*

(b) *if these values are indeed admissible, and hence a power distribution $d\mu(\theta)$ exists that is consistent with (1), then how can we describe* all *other poosible power distributions which are consistent with (1)?*

The theory in Section III gives answers to both, albeit not analytical ones. The case of planar, or even spatially distributed, arrays with several elements and an arbitrary geometry does not present any additional conceptual difficulty. We will revisit the above plus one additional numerical example of such an array after we develop the needed theory.

C. Sensor arrays – multidimensional case

We focus on two-dimensional (2D) examples. The general case of three or higher dimensions is quite similar, as is the case of near-field excitation. Most of the difficulties in dealing with dimensions higher than one, become already apparent in the 2D-case—and manageable in the same way.

Consider the spatial correlation matrix of a two-dimensional array (see [35, page 50]), i.e., neglecting temporal correlations. If the source of impinging (narrow bandwidth, stationary, etc.) waves is distributed across a sector in the sky parametrized by, say, Euler angles $(\theta_1, \theta_2) \in \mathcal{S}$, then the correlation of signals at the k^{th} and the ℓ^{th} sensor locations would produce

$$R_{k,\ell} = E\{u_k \bar{u}_\ell\} = \int_{\mathcal{S}} g_{k,\ell}(\theta_1, \theta_2) d\mu(\theta_1, \theta_2). \tag{2}$$

The integration kernels $g_{k,\ell}(\theta_1, \theta_2)$ encapsulate the relative attenuation and phase difference at the two sensor locations of a signal originating in the direction $\theta := (\theta_1, \theta_2)$. Equation (2) applies equally well when θ parametrizes a 2D-power planar distribution of nearfield sources. There is no essential difference in higher dimensional cases. The parameter θ may be thought of as belonging to $\mathbb{R}^3$, etc. and (2) thought of as a volume, etc. integral.

The same questions raised before are again central:

(a) *Given values $R_{k,\ell}$ ($k, \ell \in \{1, 2, \ldots, N\}$), how can we tell that they are consistent with the assumption of being correlation samples*?

(b) *If the answer to part (a) is yes, then what are all consistent 2D-power distributions $d\mu$*?

It was Brad Dickinson in 1980 [18] who in fact pointed out that, even if sampling takes place on a rectangular grid, positive definiteness of $[R_{k,\ell}]_{k,\ell=1}^N$ is not sufficient (see also Rudin [50]). This remark sparked considerable amount of interest in the subject and, in particular, motivated the work by Lang and McClellan [41]. Their approach was to seek a "rational" maximum entropy distribution—the idea being that if there is a distribution at all which is consistent with the data, then a maximum entropy one ought to exist as well. Their conclusion was that for 2D and beyond, this is not the case. But, as we will see in the sequel, the exponential family allows completion of their program.

III. Main results

We present two main results (Theorems 2 and 3) that provide the means to determine whether a given set of moments is admissible, and if so, to describe *all* positive measures which are consistent with the moments. Theorem 2 is essentially limited to the 1D case, and only extends to 2D under a strong assumption on the integration kernels. Theorem 3 on the other hand, applies to multivariable distributions. Notation and much of the development proceeds in parallel. We specialize to 1D or 2D only when needed.

We begin with a closed interval[1] in $\mathbb{R}^m$ ($m \in \mathbb{N}$) which we denote by $\mathcal{S}$. For notational convenience we will always take the end points to be 0 and 1 respectively, as in $\mathcal{S} = [0,1] \times [0,1]$ for the 2D case. We denote by $\mathcal{C}$ the space of real-valued, twice differentiable, (scalar) functions on $\mathcal{S}$. We denote by $\mathcal{G}$ a $(n+1)$-*dimensional subspace of* $\mathcal{C}$ generated by a set of basis functions $\{g_0, g_1, \ldots, g_n\}$. We specify that $g_0 \equiv 1$ (though it suffices to assume that $\mathcal{G}$ has a positive element). The vector-valued function

$$G := \begin{bmatrix} g_0 & g_1 & \ldots & g_n \end{bmatrix}'$$

where $'$ denotes "the transpose of", defines

$$\mathfrak{G} := \{G(\theta) \ : \ \theta \in \mathcal{S}\} \subset \mathbb{R}^{n+1}$$

as a curve, a surface, etc., depending on the dimensionality of $\mathcal{S}$. The set $\mathfrak{G}$ is often referred to as the "array manifold" in the signal processing literature. The *conic closed convex hull of* $\mathfrak{G}$ is denoted by

$$\mathfrak{K}(\mathfrak{G}) := \{R \ : \ R = \sum_{i=1}^{N} G(\theta_i) m_i, \forall \theta_i \in \mathcal{S}, m_i > 0\},$$

and its *dual cone* is denoted by

$$\mathfrak{K}_+^*(\mathfrak{G}) := \{\lambda \in \mathbb{R}^{n+1} \ : \ \langle \lambda, R \rangle \geq 0, \forall R \in \mathfrak{K}(\mathfrak{G})\}.$$

Unless we intend to emphasize the role of $\mathfrak{G}$, we will use the more compact notation $\mathfrak{K}$ and $\mathfrak{K}_+^*$ respectively, dropping the argument. We will also assume throughout that $\mathfrak{K}_+^*$ is not empty or, equivalently, that $\mathfrak{K}$ is not the entire space.

The dual cone $\mathfrak{K}_+^*$ represents the cone of all vectors λ in the dual space $(\mathbb{R}^{n+1})^*$ which form an acute or a right angle with any vector of $\mathfrak{K}$ (see Figure 7, also Figures 2 and 5). The dual space $(\mathbb{R}^{n+1})^*$ can be identified with the space where the R's live—though will always be drawn separately as in Figure 7). Then, $\langle \cdot, \cdot \rangle$ can be thought of as the standard inner product.

The dual cone $\mathfrak{K}_+^*$ can also be identified with the cone of nonnegative elements of $\mathcal{G}$ (see [39, page 14, Theorem 3.3]). The proof simply points to the fact that $\lambda \in \mathfrak{K}_+^*$ defines a support half-space of $\mathfrak{K}$, $\mathcal{H}_\lambda := \{R \ : \ \langle \lambda, R \rangle \geq 0\}$, which contains $\mathfrak{G}$. So $\langle \lambda, G \rangle \geq 0$ on $\mathcal{S}$. Conversely, if the hyperplane $\{R \ : \ \langle \lambda, R \rangle = 0\}$ cuts $\mathfrak{G}$ then it cannot be a support hyperplane for $\mathfrak{K}$. The argument holds in any dimension. It is also easy to see that λ is an interior point of $\mathfrak{K}_+^*$ if and only if $\langle \lambda, G(\theta) \rangle$ is strictly positive on $\mathcal{S}$. The standing assumption that $g_0 = 1$ implies that $\mathfrak{K}_+^*$ has at least one interior point.

The following fundamental result characterizes $\mathfrak{K}$.

Theorem 1: [39, Theorem 3.4] The cone $\mathfrak{K}(\mathfrak{G})$ is the set of points R such that

$$R = \int_{\mathcal{S}} G(\theta) d\mu(\theta), \tag{3}$$

where $d\mu$ is non-negative measure on $\mathcal{S}$.

Proof: The proof given in [39, page 15] is for $\mathcal{S} \subset \mathbb{R}$, but applies almost verbatim to any dimension. Briefly, the set of R's given by (3) is a convex conic set containing $\mathfrak{G}$ (since all points in $\mathfrak{G}$ can be obtained with a suitable choice of a singular point measure $d\mu$). The first step is to prove that this set is closed, and hence that it contains $\mathfrak{K}(\mathfrak{G})$. Consider a limit point R of a sequence R_i ($i = 1, 2, 3, \ldots$) and corresponding measures $d\mu_i$ such that $R_i = \int_{\mathcal{S}} G d\mu_i$. Using the fact that $\mathfrak{K}_+^*$ has an interior point λ, it can be shown that the sequence $\{d\mu_i \ : \ i = 1, 2, \ldots\}$ is uniformly bounded (cf. [39, page 15]). Krein and Nudelman invoke Helly's selection theorem (see [39, page 15]) to assertain that the sequence has a weak* limit $d\mu$, bounded and positive, which satisfies $R = \int_{\mathcal{S}} G d\mu$. The same is true in complete generality for positive Borel measures with support on the compact set $\mathcal{S} \subset \mathbb{R}^k$ ($k > 1$). By Riesz' representation theorem such measures make up the dual of the space of continuous functions on $\mathcal{S}$. The weak* compactness of a bounded set gives the required conclusion (see [51, pages 67 and 68]. Hence the set of R's is closed and contains $\mathfrak{K}(\mathfrak{G})$.

The argument for the converse given in [39, page 15] holds verbatim for $\mathbb{R}^k$. ∎

A non-negative measure $d\mu(\theta)$ on $\mathcal{S}$ can be thought of as a mass distribution on $\mathfrak{G}$. The set of all non-negative Borel measures on $\mathcal{S}$ is itself a closed convex cone and will be denoted by $\mathcal{M}$. The moment problem amounts to inferring properties of $d\mu$ from a vector of moments

$$R := \int_{\theta \in \mathcal{S}} G(\theta) d\mu(\theta) \in \mathbb{R}^{n+1}. \tag{4}$$

In particular, given $R \in \mathbb{R}^{n+1}$ the key questions are:

(i) does there exist $d\mu \in \mathcal{M}$ for which (4) holds?
(ii) if yes, then what all such $d\mu$'s that satisfy (4)?

Both, our assumption that $g_0 \equiv 1$ on $\mathcal{S}$ as well as our choice to work with real-valued functions on $\mathcal{S}$ can be relaxed at the expense of a more complicated notation.

Answering (i) is equivalent to determining whether a given R belongs to $\mathfrak{K}(\mathfrak{G})$ (Theorem 1 above), or altenatively, whether the functional

$$\mathfrak{C}_R \ : \ \mathfrak{K}_+^* \to \mathbb{R} \ : \ \lambda \mapsto \langle \lambda, R \rangle$$

[1]With superficial modifications it is also possible to deal with more general cases (see Section VI-A and Remark 7) where $\mathcal{S}$ is the semi-infinite interval of the nonegative reals $\mathbb{R}_+$, the reals $\mathbb{R}$, $\mathbb{R}_+^2$, etc., by attaching points at ∞ in each case, cf. [39], or with $\mathcal{S}$ being a discrete set of finitely many points $\{0, 1, \ldots, k\}$.

is nonnegative[2] (again by virtue of Theorem 1). In general, neither condition is easy.

In the classical theory of moments it is often the case that any positive element in $\mathfrak{K}_+^*$ (identified with a positive element $\sum \lambda_i g_i \in \mathcal{G}$), can be represented as a "sum of squares" (e.g., see [39, Chapter III]). In this case[3] the value of $\mathfrak{C}_R$ on positive elements can be determined via a quadratic form (having typically a Pick, Toeplitz, or Hankel structure). Naturally, positivity of $\mathfrak{C}_R$ in all such cases can be assertained with relative ease. This is not to be expected in the generality sought in the present paper.

Below we provide a way to test whether a given vector R of moments belongs to the interior $\text{int}(\mathfrak{K})$ of the cone $\mathfrak{K}$, in considerable generality. This is accomplished by constructing explicitely a $d\mu \in \mathcal{M}$ which satisfies (4). The particular $d\mu$ is obtained by integrating a certain differential equation. When $R \notin \text{int}(\mathfrak{K}(\mathfrak{G}))$ the differential equation diverges. When it converges, the limit point provides a set of parameters that identify a suitable $d\mu$.

We exploit two special families of measures,

$$\mathcal{M}_{\text{rat}} := \{\frac{1}{< \lambda, G(\theta) >} d\theta \ : \ \text{with } \lambda \in \text{int}(\mathfrak{K}_+^*)\}$$

and

$$\mathcal{M}_{\text{exp}} := \{e^{-<\lambda, G(\theta)>} d\theta \ : \ \text{with } \lambda \in \text{int}(\mathfrak{K}^*)\}.$$

The elements of $\mathcal{M}_{\text{rat}}$ are in bijective correspondence with interior points in $\mathfrak{K}(\mathfrak{G})$ in the 1D case and, with some strong assumption, in the 2D as well. The elements of $\mathcal{M}_{\text{exp}}$ are in bijective correspondence with interior points in $\mathfrak{K}(\mathfrak{G})$ in complete generality. Similar facts hold true for

$$\mathcal{M}_{\text{rat},\Psi} := \{\frac{\Psi(\theta)}{< \lambda, G(\theta) >} d\theta \ : \ \text{with } \lambda \in \text{int}(\mathfrak{K}_+^*)\}$$

and

$$\mathcal{M}_{\text{exp},\Psi} := \{\Psi(\theta) e^{-<\lambda, G(\theta)>} d\theta \ : \ \text{with } \lambda \in \text{int}(\mathfrak{K}^*)\}$$

where Ψ is an arbitrary but fixed positive function in $\mathcal{C}$. The last two families allow characterization of *all* solutions consistent with a given $R \in \text{int}(\mathfrak{K})$ in a non-classical fashion.

These families represent extrema of certain logarithmic entropy-like functionals subject to the moment constraint (4). For the case of $\mathcal{M}_{\text{rat}}$ and $\mathcal{M}_{\text{exp}}$, the relevant functionals are $\int \log(\frac{d\mu}{d\theta}) d\theta$ and $\int \log(\frac{d\mu}{d\theta}) d\mu(\theta)$, respectively. In the more general case of $\mathcal{M}_{\text{rat},\Psi}$ and $\mathcal{M}_{\text{exp},\Psi}$, the functionals can be written in the Kullback-Leibler form:

$$S(d\mu_0 || d\mu_1) := \int \log\left(\frac{d\mu_0}{d\mu_1}\right) d\mu_0,$$

with $d\mu_\ell = \Psi d\theta$ and $d\mu_{(\ell+1)_{\text{mod}2}} \equiv d\mu$, respectively for $\ell = 0, 1$. The fact that they provide minimizers to entropy functionals is not being exploited at all in the present analysis. Instead, we consider directly the solvability of (4), and analyze it with a continuation method.

The search for extremals of entropy for identifying spectra has a long and well known history. In the context of our present analysis, the work of Lang and McClellan [41] is particularly relevant. In this early work, the authors sought measures within $\mathcal{M}_{\text{rat}}$ for problems with multi-dimensional support. They soon discovered that this family is not "rich" enough, and then suggested replacing the multivariable support of the measures with a finite net of points—effectively restricting attention to 0-dimensional spectral support. Interest was rekindled in recent years by Byrnes, Gusev, and Lindquist [12] with a clever constructive solution of the rational covariance extension problem of [23], [24], based on minimizers of relative entropy. Attention was then drawn to $\mathcal{M}_{\text{rat},\Psi}$ and to the effect of selecting Ψ (first as a positive rational function of a given degree in [12], [10], [11] and then, in increasing generality, in [13], [14], [31]).

The theory is developed in parallel for the rational and exponential families since most of the steps are similar. We first state the more restricted version which is based on the rational family and then the more general one based on the exponential family.

Theorem 2: Consider $\mathcal{S}, \mathcal{C}, \mathcal{G}, G$, a positive element $\Psi \in \mathcal{C}$, and a vector $R_1 \in \mathbb{R}^{n+1}$. Assume that the set $\mathcal{S}$ is either a closed interval in $\mathbb{R}$ or in $\mathbb{R}^2$. In case $\mathcal{S}$ is an interval in $\mathbb{R}^2$, assume that G is doubly periodic. Consider the differential equation

$$\frac{d\lambda(t)}{dt} = f_{\text{rat}}(\lambda(t)), \tag{5}$$

where

$$\begin{aligned} f_{\text{rat}}(\lambda) &= -M(\lambda)^{-1}(R_1 - \int_{\mathcal{S}} G(\theta) \frac{\Psi(\theta)}{\langle \lambda, G(\theta) \rangle} d\theta), \\ M(\lambda) &= \int_{\mathcal{S}} \left(G(\theta) \frac{\Psi(\theta) d\theta}{\langle \lambda(t), G(\theta) \rangle^2} G(\theta)' \right) d\theta, \end{aligned} \tag{6}$$

and $\lambda(0) = \lambda_0 \in \text{int}(\mathfrak{K}_+^*)$ (e.g., $\lambda_0 = (1, 0, \ldots, 0)'$). If $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$, then as $t \to \infty$ the solution $\lambda(t)$ of (5) tends to a limit $\lambda_1 \in \text{int}(\mathfrak{K}_+^*)$ that satisfies

$$R_1 = \int_{\mathcal{S}} G(\theta) \frac{\Psi(\theta)}{\langle \lambda_1, G(\theta) \rangle} d\theta. \tag{7}$$

Moreover:
(i) the trajectory $\{\lambda(t) : t \in [0, \infty)\}$ remains in $\text{int}(\mathfrak{K}_+^*)$,
(ii) the limit point λ_1 is the unique solution of (7) in $\mathfrak{K}_+^*$ and
(iii) the convergence $\lambda(t) \to \lambda_1$ is exponential with a Lyapunov function

$$V(\lambda) = \| R_1 - \int_{\mathcal{S}} G(\theta) \frac{\Psi(\theta)}{\langle \lambda, G(\theta) \rangle} d\theta \|^2$$

satisfying

$$\frac{dV(\lambda(t))}{dt} = -2V(\lambda(t))$$

along trajectories of (5).

[2]A functional $\mathfrak{C}_R$ is said to be nonnegative if it takes $\mathfrak{K}_+^*$ into $\mathbb{R}_+$, and it is said to be positive if it is nonnegative and does not vanish on nonzero elements of $\mathfrak{K}_+^*$.

[3]Similarly, when $\mathcal{G}$ is invariant under a "backward shift" operator, see [26], [27]; see also [25, page 786] for a negative result.

Conversely, if $R_1 \notin \text{int}(\mathfrak{K}(\mathfrak{G}))$ then $\|\lambda(t)\| \to \infty$.

A completely analogous statement holds true for the exponential family, with the added advantage that $\mathcal{S}$ can now be an interval in $\mathbb{R}^k$ ($k \geq 1$).

Theorem 3: Consider $\mathcal{S}, \mathcal{C}, \mathcal{G}, G$, a positive element $\Psi \in \mathcal{C}$, and a vector $R_1 \in \mathbb{R}^{n+1}$. The set $\mathcal{S}$ is assumed to be a closed interval in $\mathbb{R}^k$ with $k \geq 1$. Consider the differential equation

$$\frac{d\lambda(t)}{dt} = f_{\exp}(\lambda(t)), \tag{8}$$

where

$$\begin{aligned} f_{\exp}(\lambda) &= -M(\lambda)^{-1}(R_1 - \int_{\mathcal{S}} G(\theta)\Psi(\theta)e^{-\langle \lambda, G(\theta)\rangle}d\theta) \\ M(\lambda) &= \int_{\mathcal{S}} \left(G(\theta)\Psi(\theta)e^{-\langle \lambda, G(\theta)\rangle}G(\theta)'\right)d\theta \end{aligned}$$

and $\lambda(0) = \lambda_0 \in \mathbb{R}^{n+1}$, e.g., $\lambda_0 = (1,\ 0,\ \ldots,\ 0)'$. If $R_1 \in \text{int}\mathfrak{K}(\mathfrak{G})$, then as $t \to \infty$ the solution $\lambda(t)$ of (8) tends to a limit $\lambda_1 \in \mathbb{R}^{n+1}$ that satisfies

$$R_1 = \int_{\mathcal{S}} G(\theta)\Psi(\theta)e^{-\langle \lambda_1, G(\theta)\rangle}d\theta. \tag{9}$$

Moreover:
(i) the trajectory $\{\lambda(t) : t \in [0,\infty)\}$ remains bounded,
(ii) the limit point λ_1 is the unique solution of (9) in $\mathbb{R}^{n+1}$, and
(iii) the convergence $\lambda(t) \to \lambda_1$ is exponential with a Lyapunov function

$$V(\lambda) = \|R_1 - \int_{\mathcal{S}} G(\theta)\Psi(\theta)e^{-\langle \lambda, G(\theta)\rangle}d\theta\|^2$$

satisfying

$$\frac{dV(\lambda(t))}{dt} = -2V(\lambda(t))$$

along trajectories of (8).
Conversely, if $R_1 \notin \text{int}\mathfrak{K}(\mathfrak{G})$ then $\|\lambda(t)\| \to \infty$.

Theorems 2 and 3 hold when $\mathcal{S}$ is a discrete set, in which case integration is to be replaced by summation over $\mathcal{S}$, and $\mathcal{C}, \mathcal{G}, G$ are interpreted accordingly.

IV. An analytic example

It is often insightful to work through a simple analytic example. We do this for the case where

$$G(\theta) = \begin{bmatrix} 1 \\ \theta \end{bmatrix} \text{ and } \mathcal{S} = [0,1],$$

in order to explain and highlight certain facts that are exploited in the proofs of Theorems 2 and 3.

Since elements in $\mathfrak{K}_+^*$ can be identified with polynomials $\lambda_0 + \lambda_1\theta$ which are nonnegative on $\mathcal{S}$, it is not hard to see that

$$\mathfrak{K}_+^* = \left\{ \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \ :\ \lambda_0 \geq 0,\ \lambda_1 \geq -\lambda_0 \right\}.$$

Then $\mathfrak{K}$, being its dual cone, is given by

$$\mathfrak{K} = \left\{ \begin{bmatrix} r_0 \\ r_1 \end{bmatrix} \ :\ r_0 \geq r_1 \geq 0 \right\}.$$

A schematic is shown in Figures 2 and 5 (with detail intent to explain the correspondence between moments and parameters in two separate cases discussed below).

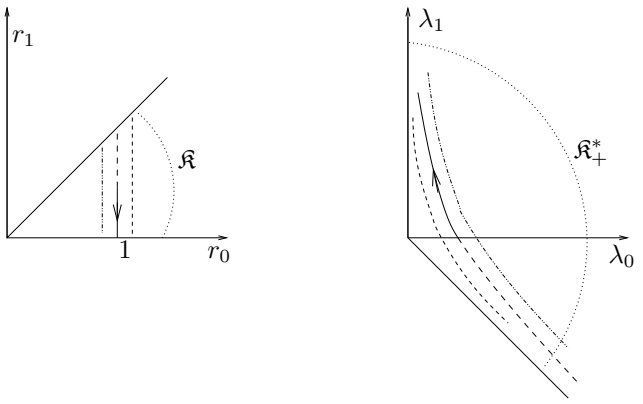

Fig. 2. Schematic of $\mathfrak{K}$ and $\mathfrak{K}_+^*$, and of the correspondence $(r_0, r_1) \leftrightarrow (\lambda_0, \lambda_1)$ for Case I.

We restrict our attention to $\Psi = 1$ and determine explicitly the correspondence

$$h : \lambda \mapsto R = \int_0^1 G(\theta)m(\theta,\lambda)d\theta$$

for the two cases, $m(\theta,\lambda) = 1/(\lambda_0 + \lambda_1\theta)$ and $m(\theta,\lambda) = e^{-(\lambda_0+\lambda_1\theta)}$. In either case we determine the values of the λ's that correspond to $r_0 = 1$ and $1 > r_1 > 0$ and explain the limiting behavior of $m(\theta,\lambda)$ as $(r_0,\ r_1)'$ is taken near the boundary of $\mathfrak{K}$.

Case I: $\mathbf{m}(\theta,\lambda) = \mathbf{1}/(\lambda_\mathbf{0} + \lambda_\mathbf{1}\theta)$

$$h : \mathfrak{K}_+^* \to \mathfrak{K} \ :\ \lambda \mapsto R = \int_0^1 \begin{bmatrix} 1 \\ \theta \end{bmatrix} \frac{1}{\lambda_0 + \lambda_1\theta}d\theta.$$

We readily compute that

$$\begin{aligned} r_0 &= \int_0^1 \frac{1}{\lambda_0 + \lambda_1\theta}d\theta = \frac{1}{\lambda_1}\log(1 + \frac{\lambda_1}{\lambda_0}),\ \text{and} \\ r_1 &= \int_0^1 \frac{\theta}{\lambda_0 + \lambda_1\theta}d\theta = -\frac{\lambda_0}{\lambda_1^2}\log(1 + \frac{\lambda_1}{\lambda_0}) + \frac{1}{\lambda_1}. \end{aligned}$$

These expressions can be rewritten in the form

$$r_0 = \frac{1}{\lambda_0}\frac{\log(1+x)}{x},\ \text{and} \tag{10}$$

$$r_1 = \frac{1}{\lambda_0}\frac{x - \log(1+x)}{x^2},\ \text{where} \tag{11}$$

$$x = \frac{\lambda_1}{\lambda_0}. \tag{12}$$

We oberve that

$$\frac{r_1}{r_0} = \frac{1}{\log(1+x)} - \frac{1}{x} \tag{13}$$

tends to $\frac{1}{2}$ as $x \to \pm 0$, it tends to 1 as $x \searrow -1$, it tends to 0 as $x \to \infty$, and has a negative derivative throughout $(0, \infty)$. Hence, if we specify that $r_1/r_0 = 1/2$ when $x = 0$, equation (13) defines a continuous and monotonically decreasing function of x in $[-1, \infty)$, see Figure 3.

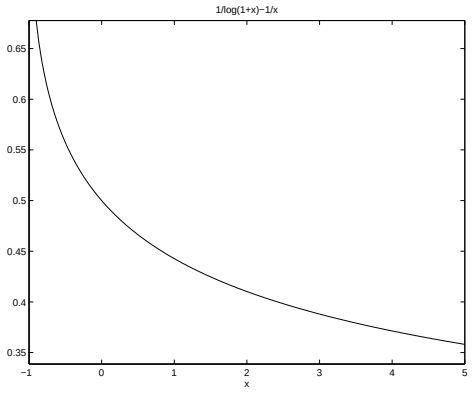

Fig. 3. r_1/r_0 vs. $x := \lambda_1/\lambda_0$

Equation (13) thus defines a bijective correspondence between the ratios $\lambda_1/\lambda_0 \in [-1, \infty)$ and $r_1/r_0 \in (0, 1]$. Then from (10) we see that the product of r_0 and λ_0 depends only on these ratios. Therefore h is a bijective map between pairs (r_0, r_1) and (λ_0, λ_1) in the interior $\mathfrak{K}$ and $\mathfrak{K}_+^*$, respectively.

A final point to be made is that when (r_0, r_1) is taken to approach the boundary of $\mathfrak{K}$ then $\|(\lambda_0,\ \lambda_1)\|$ grows unbounded. To see this, for simplicity, we take $r_0 = 1$: From (10), we have that

$$\begin{aligned} \lambda_0 &= \frac{\log(1+x)}{x} \text{ while} \\ \lambda_1 &= \log(1+x). \end{aligned}$$

Hence

$$\begin{aligned} \frac{r_1}{r_0} \nearrow 1 \quad &\Rightarrow \quad x \searrow -1 \Rightarrow \lambda_0 \to +\infty, \text{ while} \\ \frac{r_1}{r_0} \searrow 0 \quad &\Rightarrow \quad x \nearrow +\infty \Rightarrow \lambda_1 \to +\infty. \end{aligned}$$

Either way, $m(\theta) = 1/(\lambda_0 + \lambda_1\theta)$ tends to a singular measure. Indeed, since $\|(\lambda_0,\ \lambda_1)\|$ grows unbounded, $m(\theta)$ tends to zero except at points where the denominator tends to vanish.

Case II: $\mathbf{m}(\theta, \lambda) = \mathbf{e}^{-(\lambda_0 + \lambda_1\theta)}$

In this case $(\lambda_0,\ \lambda_1) \in \mathbb{R}^2$ with no restrictions and

$$h : \mathbb{R}^2 \to \mathfrak{K} \ : \ \lambda \mapsto R = \int_0^1 \begin{bmatrix} 1 \\ \theta \end{bmatrix} e^{-(\lambda_0 + \lambda_1\theta)} d\theta.$$

We readily compute that

$$\begin{aligned} r_0 &= e^{-\lambda_0} \frac{1 - e^{-\lambda_1}}{\lambda_1}, \text{ and} \\ r_1 &= e^{-\lambda_0} \frac{1 - e^{-\lambda_1} - \lambda_1 e^{-\lambda_1}}{\lambda_1^2}, \end{aligned}$$

which shows that this time the ratio

$$\frac{r_1}{r_0} = \frac{1 - e^{-\lambda_1} - \lambda_1 e^{-\lambda_1}}{\lambda_1(1 - e^{-\lambda_1})}, \tag{14}$$

can take any value in $(0, 1)$ for a corresponding *unique* value of λ_1. A plot of the r_1/r_0 vs. λ_1 is shown in Figure (4). Then $\lambda_0 = \log(\frac{1-e^{-\lambda_1}}{r_0\lambda_1})$ Thus, there is a bijective correspondence between $(r_0,\ r_1)$ in the interior of $\mathfrak{K}$ and points $(\lambda_0,\ \lambda_1) \in \mathbb{R}^2$.

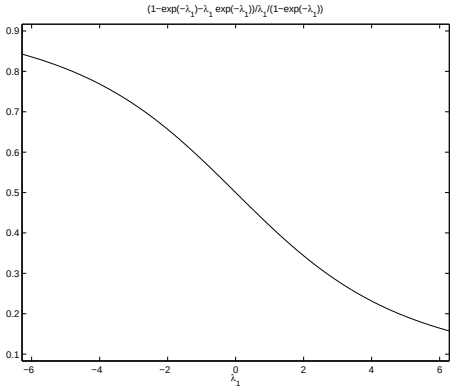

Fig. 4. r_1/r_0 vs. λ_1

Yet again, when $(r_1,\ r_2)$ is taken to approach the boundary of $\mathfrak{K}$ (i.e., either $r_1/r_0 \searrow 0$ or $r_1/r_0 \nearrow 1$), $\|(\lambda_0,\ \lambda_1)\|$ grows unbounded. This can be seen from Figure 4 and can easily verified directly. However, an interesting fact which will be exploited in the proof of Theorem 3, is that very much like in the previous case of the rational family, $m(\theta)$ once again tends to a singular measure *with singularities at the roots of an element in* $\mathfrak{K}_+^*$. The difference with the earlier case is that this time, the curve $\lambda_0 = \log(\frac{1-e^{-\lambda_1}}{r_0\lambda_1})$ may lie outside $\mathfrak{K}_+^*$ depending on the value of r_0. This is sketched in Figure 5. Yet, as $\lambda_1 \to \pm\infty$, it again "lines up"

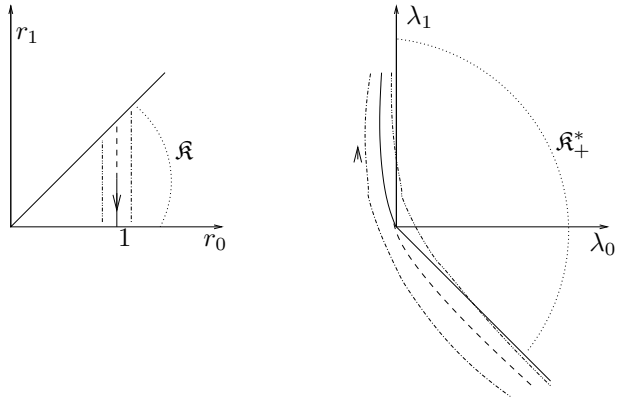

Fig. 5. Schematic of $\mathfrak{K}$ and $\mathfrak{K}_+^*$, and of the correspondence $(r_0, r_1) \leftrightarrow (\lambda_0, \lambda_1)$ for Case II.

with the boundary of $\mathfrak{K}_+^*$. Indeed,

$$\frac{r_1}{r_0} \nearrow 1 \quad \Rightarrow \quad \lambda_1 \searrow -\infty \Rightarrow \frac{\lambda_0}{\lambda_1} \to -1, \text{ while}$$
$$\frac{r_1}{r_0} \searrow 0 \quad \Rightarrow \quad \lambda_1 \nearrow +\infty \Rightarrow \frac{\lambda_0}{\lambda_1} \to 0.$$

It is quite revealing to consider how this works: if for instance, $r_0 = 1$ and $1 > r_1 \sim 1$, then $-(\lambda_0 + \lambda_1\theta)$ becomes large and negative over most of $[0, 1]$, except near 1. In a small neighborhood of 1 it becomes "positive enough" so that $\int_0^1 e^{-(\lambda_0+\lambda_1\theta)}d\theta = 1$. Figure 6 displays $-(\lambda_0 + \lambda_1\theta)$ and $e^{-(\lambda_0+\lambda_1\theta)}$ as a function of θ, for one such set of values.

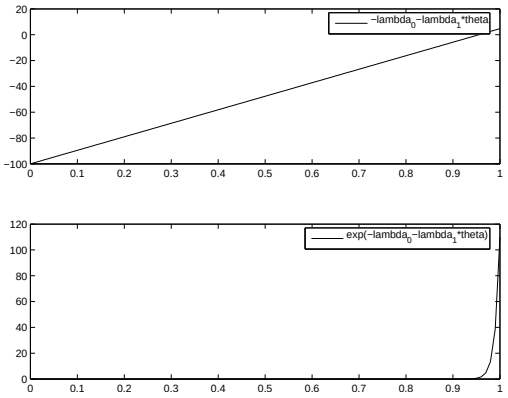

Fig. 6. $-(\lambda_0 + \lambda_1\theta)$ and $e^{-(\lambda_0+\lambda_1\theta)}$ as a function of θ

Remark 4: As it will become apparent in the next section, when the vector of moments R is taken to approach the boundary of $\mathfrak{K}$, both, the rational density $\Psi/\langle\lambda, G\rangle$ as well as the exponential one $\Psi e^{-\langle\lambda,G\rangle}$ in Theorems 2 and 3, tend to become singular with singularities at the roots of a boundary element of $\mathfrak{K}_+^*$. At times this observation may be useful in determining the maximum number of singularities ("spectral lines") that such families of density functions are capable of reproducing.

V. Proofs of Theorems 2 and 3.

The arguments for both theorems are similar and thus, for the most part, proceed in parallel. The key idea, is to study and compute solutions for the nonlinear equations (7) and (9) via a one-parameter imbedding (homotopy). This is a standard technique in nonlinear analysis (see [46] and [38]) and amounts to introducing a one-parameter family of problems, from one which is readily solvable to the one of interest. This we do below by introducing a path $R_\rho = (1-\rho)R_0 + \rho R_1$, $\rho \in [0, 1]$, from a convenient choice R_0 for a vector of moments to the given R_1. We then consider the family of equations

$$R_\rho = \int_S G(\theta)d\mu_\rho(\theta), \text{ with } \rho \in [0, 1] \tag{15}$$

in order to link a known $d\mu_0$ to the sought-after solution $d\mu_1$, with $d\mu_\rho$ in an appropriate family parametrized by a vector $\lambda(\rho)$. A schematic is shown in Figure 7. At the end, we re-scale the homotopy variable ρ, replacing it with a new variable $t \in [0, \infty)$, so as to bring the differential equation which connects the family of respective solutions into the feedback form given in the two theorems.

First, a note regarding notation: both families $\mathcal{M}_{\text{rat}}$ and $\mathcal{M}_{\text{exp}}$ consist of absolutely continuous positive measures $d\mu$ with bounded derivative, parametrized by λ. For both, we denote by $m(\theta) = \dot{\mu}(\theta)$ the corresponding "density" functions and, whenever we want to emphasize their dependence on λ we use the notation $m(\theta, \lambda)$. At places we need to indicate an additional dependence on a homotopy variable ρ, which we then introduce as a subscript, i.e., as in $m_\rho(\theta, \lambda)$.

We begin with a given vector of moments $R_1 \in \mathbb{R}^{n+1}$ and we are interested in the solvability of the nonlinear equation

$$R_1 = \int_{\mathcal{S}} G(\theta)m(\theta, \lambda)d\theta \tag{16}$$

with $m(\theta, \lambda)d\theta$ in one of $\mathcal{M}_{\text{rat}}$ or $\mathcal{M}_{\text{exp}}$. We know a "starting" pair, $(R_0, m(\theta, \lambda_0))$ with $R_0 \in \mathfrak{K}$ and $m(\theta, \lambda_0)d\theta$ in the appropriate family. Simply take $m(\theta, \lambda_0) = \frac{\Psi}{\langle\lambda_0,G\rangle}$ in $\mathcal{M}_{\text{rat}}$ or $m(\theta, \lambda_0) = \Psi e^{-\langle\lambda_0,G\rangle}$ in $\mathcal{M}_{\text{exp}}$, with $\lambda_0 = (1, 0, \ldots, 0)$ for either, and compute the corresponding vector of moments using

$$R_0 = \int_{\mathcal{S}} G(\theta)m(\theta, \lambda_0)d\theta. \tag{17}$$

Our plan is to study the one-parameter homotopy

$$H(\rho, \lambda) := \int_{\mathcal{S}} G(\theta)m_\rho(\theta, \lambda)d\theta - R_\rho = 0 \tag{18}$$

where

$$R_\rho = (1-\rho)R_0 + \rho R_1 \text{ and } \rho \in [0, 1], \tag{19}$$

and trace the family of respective solutions. Evidently, $H(1, \lambda) = 0$ is equivalent to (16) for which a solution is being sought, while $H(0, \lambda) = 0$ is equivalent to (17) for which a solution is available.

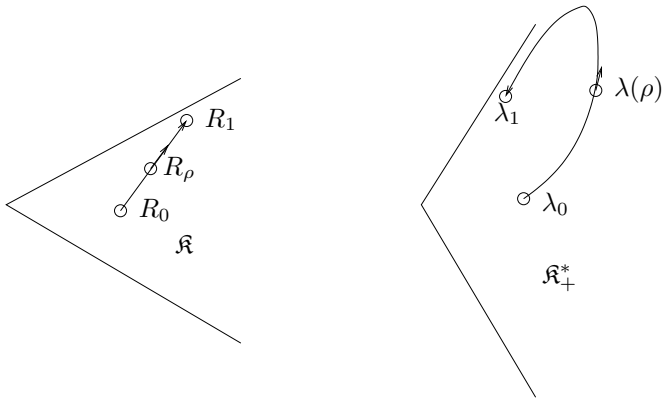

Fig. 7. Schematic of the dual cones $\mathfrak{K}$ and $\mathfrak{K}_+^*$, and of R_ρ and λ_ρ, for $\rho \in [0, 1]$.

If the equation $H(\rho, \lambda) = 0$ has a solution $\lambda(\rho)$ such that

$$\left.\frac{\partial H(\rho, \lambda)}{\partial \lambda}\right|_{\rho,\lambda(\rho)}$$

is non-singular for $\rho \in [0,1]$, then it follows from the implicit function theorem that $\lambda(\rho)$ satisfies the differential equation

$$\frac{d\lambda(\rho)}{d\rho} = -\left(\frac{\partial H(\rho,\lambda)}{\partial \lambda}\right)^{-1} \frac{\partial H(\rho,\lambda)}{\partial \rho}, \text{ with } \lambda(0)=\lambda_0. \tag{20}$$

Conversely, if (20) has a solution $\lambda(\rho)$ for $\rho \in [0,1]$ then $dH(\rho,\lambda(\rho))/d\rho = 0$ for $\rho \in [0,1]$. Therefore $H(0,\lambda(0)) = H(1,\lambda(1)) = 0$, and $\lambda(1)$ gives rise to an $m(\theta,\lambda)$ which satisfies (16).

Since $\partial H(\rho,\lambda)/\partial\rho = R_0 - R_1$, equation (20) becomes

$$\frac{d\lambda(\rho)}{d\rho} = -\left(\frac{\partial H(\rho,\lambda)}{\partial \lambda}\right)^{-1} (R_0 - R_1), \text{ with } \lambda(0)=\lambda_0. \tag{21}$$

On the other hand, we readily compute that the partial derivative with respect to λ is

$$\frac{\partial H(\rho,\lambda)}{\partial \lambda} = -\int_{\mathcal{S}} G(\theta)\frac{\Psi(\theta)}{\langle \lambda, G(\theta)\rangle^2} G(\theta)' d\theta =: -M_{\rm rat}(\lambda) \tag{22}$$

in the case of Theorem 2, and

$$\frac{\partial H(\rho,\lambda)}{\partial \lambda} = -\int_{\mathcal{S}} G(\theta)\Psi(\theta) e^{-\langle \lambda, G(\theta)\rangle} G(\theta)' d\theta =: -M_{\rm exp}(\lambda) \tag{23}$$

for Theorem 3. We will show that $M_{\rm exp}$ is bounded and invertible along trajectories of (20) for all $\rho \in [0,1]$, if and only if $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$. The same is true for $M_{\rm rat}$ provided $\mathcal{S}$ is of dimension 1, or of dimension 2 and G is periodic in its arguments. When that happens, R_ρ can be computed either through (19) or, through (15) using $d\mu_\rho = m(\theta,\lambda(\rho))d\theta$ with $\lambda(\rho)$ obtained via (21).

Using (19) we have that

$$R_1 - R_0 = \frac{1}{1-\rho}(R_1 - R_\rho).$$

Now, by changing variables in (20) via

$$\rho = 1 - e^{-t}$$

and substituting $R_1 - R_0$, we obtain

$$\begin{aligned}
\frac{d\lambda(t)}{dt} &= -\left(\frac{\partial H(\rho(t),\lambda)}{\partial \lambda}\right)^{-1} \frac{\partial H(\rho(t),\lambda)}{\partial \rho}\frac{d\rho}{dt} \\
&= -\left(\frac{\partial H(\rho(t),\lambda)}{\partial \lambda}\right)^{-1} (R_0 - R_1)(1-\rho) \\
&= -\left(\frac{\partial H(\rho(t),\lambda)}{\partial \lambda}\right)^{-1} (R_\rho - R_1)
\end{aligned} \tag{24}$$

(where $d\lambda(t)/dt$ denotes $d\lambda(\rho(t))/dt$, following a common simplification of notation). In view of (22), (23), and (15) the equivalence between (20) and the differential equations in the two theorems, namely (5) and (8), is now complete. We only need to prove the claim that $M_{\rm rat}$ and $M_{\rm exp}$ remain invertible along a trajectories of (20) for $\rho \in [0,1]$ if and only if $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$, under the stated assumptions. We first establish this in the context of Theorem 2.

Proposition 5: Let $\mathcal{S}, G$ be as in Theorem 2 and λ_0, R_0 be as above. The integral $\int_{\mathcal{S}} G(\theta)\frac{\Psi(\theta)}{\langle \lambda, G(\theta)\rangle^2} G(\theta)' d\theta = M_{\rm rat}(\lambda)$ remains bounded and nonsingular along the trajectory of

$$\frac{d\lambda}{d\rho} = -M_{\rm rat}(\lambda)^{-1}(R_1 - R_0) \tag{25}$$

for $\rho \in [0,1]$ if and only if $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$.

Proof: Since $\mathcal{S}$ is closed and $\langle \lambda, G\rangle$ is continuous on $\mathcal{S}$, as long as $\lambda \in \text{int}(\mathfrak{K}_+^*)$, the function $\langle \lambda, G\rangle$ is bounded away from 0 and $G\frac{\Psi}{\langle \lambda, G\rangle^2} G'$ is integrable on $\mathcal{S}$. The integral is also nonsingular because $\Psi > 0$ and the entries of G are linearly independent on $\mathcal{S}$. All this is certainly true for λ_0 and hence the differential equation can be integrated on a maximal interval $[0,\epsilon)$. Throughout, $\lambda(\rho)$ must remain in $\text{int}(\mathfrak{K}_+^*)$ for otherwise $G\Psi G'/\langle \lambda, G\rangle^2$ is not integrable.

If $\epsilon > 1$ then $dH(\rho,\lambda(\rho))/d\rho = 0$ for $\rho \in [0,1]$, and hence,

$$m(\theta) := \Psi(\theta)/\langle \lambda(1), G(\theta)\rangle \tag{26}$$

satisfies (16). This is a positive function and hence $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$.

Let now $\epsilon \leq 1$ and compute R_ρ using

$$R_\rho = \int_{\mathcal{S}} G(\theta)\frac{\Psi(\theta)}{\langle \lambda(\rho), G(\theta)\rangle} d\theta \tag{27}$$

on the trajectory $\lambda(\rho)$ of (25) for $\rho \in [0,\epsilon)$. It follows that

$$\begin{aligned}
\frac{dR_\rho}{d\rho} &= \frac{\partial}{\partial \lambda}\left(\int_{\mathcal{S}} G(\theta)\frac{\Psi(\theta)}{\langle \lambda(\rho), G(\theta)\rangle} d\theta\right)\frac{d\lambda}{d\rho} \\
&= R_1 - R_0
\end{aligned}$$

throughout $[0,\epsilon)$, in agreement with $R_\rho = (1-\rho)R_0 + \rho R_1$. We assume R_1 belongs to $\text{int}(\mathfrak{K}(\mathfrak{G}))$ and derive a contradiction.

If $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$, then so does R_ϵ (seen as the limit of R_ρ in (27) when $\rho \to \epsilon$). Yet, by our assumption that $[0,\epsilon)$ is a maximal interval for which $\lambda(\rho) \in \text{int}(\mathfrak{K}_+^*)$, as $\rho \to \epsilon$ either $\|\lambda(\rho)\| \to \infty$ or $\lambda(\rho)$ tends to a boundary point of $\mathfrak{K}_+^*$. We first show that in fact it is always the case that $\|\lambda(\rho)\| \to \infty$. If we momentarily assume that $\lambda(\rho)$ remains bounded and if λ_c denotes a limit point of $\lambda(\rho)$ as $\rho \to \epsilon$, then $\langle \lambda_c, G\rangle$ vanishes at some point on $\mathcal{S}$. It is here that we need the continuity properties of G.

When $\mathcal{S}$ is 1-dimensional, vanishing of $\langle \lambda_c, G\rangle$ at some point in $\mathcal{S}$ implies that $\frac{\Psi}{\langle \lambda_c, G\rangle}$ is not integrable. A similar conclusion can be drawn in the case where $\mathcal{S}$ is 2-dimensional, but in this case we need G to be periodic (and twice differentiable). Then, $\mathcal{S}$ can be thought of as the torus and $\langle \lambda_c, G\rangle$ must have a double root. Consequently, $\frac{\Psi}{\langle \lambda_c, G\rangle}$ again fails to be integrable.

Lack of integrability in either case contradicts the fact that $\int_{\mathcal{S}} G(\theta)\frac{\Psi(\theta)}{\langle \lambda(\rho), G(\theta)\rangle} d\theta$ lies in the interval between R_0 and R_1 for $\rho \in [0,\epsilon)$. This is because the first entry of last integral is simply $\int_{\mathcal{S}} 1 \cdot \frac{\Psi(\theta)}{\langle \lambda(\rho), G(\theta)\rangle} d\theta$ and fails to be bounded. We conclude that the only possibility is that $\lim_{\rho\to\epsilon} \|\lambda(\rho)\| = \infty$.

We now show that if $\lim_{\rho\to\epsilon}\|\lambda(\rho)\|=\infty$, the functional

$$\mathfrak{C}_{R_\epsilon}\ :\ \mathfrak{K}_+^*\to\mathbb{R}\ :\ \lambda\mapsto\langle\lambda,R_\epsilon\rangle$$

is not strictly positive, which contradicts the assumption that $R_\epsilon\in\text{int}(\mathfrak{K}(\mathfrak{G}))$. This last claim follows readily from the fact that

$$\mathfrak{C}_{R\rho}(\lambda(\rho))=\langle\lambda(\rho),\int_{\mathcal{S}}G(\theta)\frac{\Psi(\theta)}{\langle\lambda(\rho),G(\theta)\rangle}d\theta\rangle\equiv 1$$

is valid throughout $[0,\epsilon)$, from $\mathfrak{C}_{R_\rho}$ being continuous, and from the assertion that $\|\lambda(\rho)\|\to\infty$. Indeed, $\mathfrak{C}_{R_\rho}(\lambda(\rho)\rangle$ being identically equal to 1 throughout $[0,\epsilon)$ implies that there is a sequence $R_{\rho_i}\to R_\epsilon$ for which the corresponding functionals $\mathfrak{C}_{R_{\rho_i}}$ take values $\frac{1}{\|\lambda_{\rho_i}\|}\to 0$ on vectors $\frac{1}{\|\lambda_{\rho_i}\|}\lambda_{\rho_i}$ of unit length. Thus, they are not bounded away from 0 and neither does their limit $\mathfrak{C}_{R_\epsilon}$. Therefore, $R_\epsilon\notin\text{int}(\mathfrak{K}(\mathfrak{G}))$, which gives the sought contradiction. ■

We proceed with an analogous statement in context of Theorem 3.

Proposition 6: Let $\mathcal{S},G,\lambda_0,R_0$ be as defined earlier. The integral $\int_{\mathcal{S}}G(\theta)\Psi(\theta)e^{-\langle\lambda,G(\theta)\rangle}G(\theta)'d\theta=M_{\exp}(\lambda)$ remains bounded and nonsingular along a trajectories of

$$\frac{d\lambda}{d\rho}=-M_{\exp}(\lambda)^{-1}(R_1-R_0)\tag{28}$$

for $\rho\in[0,1]$ if and only if $R_1\in\text{int}(\mathfrak{K}(\mathfrak{G}))$.

Proof: For any value of λ, $e^{-\langle\lambda,G\rangle}$ is positive and bounded throughout $\mathcal{S}$. Hence, $G\Psi e^{-\langle\lambda,G\rangle}G'$ is integrable, and because the elements of G are linearly independent, the integral is an invertible matrix. It follows that (28) can be integrated on a maximal interval $[0,\epsilon)$.

If $\epsilon>1$, we see as before that $dH(\rho,\lambda(\rho))/d\rho=0$ for $\rho\in[0,1]$, then $H(0,\lambda(0))=H(1,\lambda(1))=0$, and hence,

$$m(\theta):=\Psi(\theta)e^{-\langle\lambda(1),G(\theta)\rangle}\tag{29}$$

satisfies (16). This again is a positive function and hence $R_1\in\text{int}(\mathfrak{K}(\mathfrak{G}))$.

Conversely, if $\epsilon\le 1$ then $\|\frac{d\lambda(\rho)}{d\rho}\|$ and $\|\lambda(\rho)\|$ increase without bound, while

$$\int_{\mathcal{S}}G(\theta)\Psi(\theta)e^{-\langle\lambda(\rho),G(\theta)\rangle}G(\theta)'d\theta\tag{30}$$

tends to become singular as $\rho\to\epsilon$. We again compute

$$R_\rho=\int_{\mathcal{S}}G(\theta)\Psi(\theta)e^{-\langle\lambda(\rho),G(\theta)\rangle}d\theta\tag{31}$$

on $[0,\epsilon)$, which is in agreement with $R_\rho=(1-\rho)R_0+\rho R_1$. We will show that

$$R_\epsilon\notin\text{int}(\mathfrak{K}(\mathfrak{G})),\tag{32}$$

which implies that $R_1\notin\text{int}(\mathfrak{K}(\mathfrak{G}))$ either, due to the convexity of $\mathfrak{K}(\mathfrak{G})$ and the fact that $R_0\in\text{int}(\mathfrak{K}(\mathfrak{G}))$.

Let κ_c be the limit of a convergent sequence $\kappa_i:=\frac{1}{\|\lambda(\rho_i)\|}\lambda(\rho_i)$, with ρ_i a suitably selected increasing sequence in $[0,\epsilon)$ tending to ϵ. That such a convergent sequence exists follows from the boundedness of $\lambda(\rho)/\|\lambda(\rho)\|$ and the fact that these are finite dimensional vectors. We claim that κ_c is a boundary point of $\mathfrak{K}_+^*$. To see this note that

$$\int_{\mathcal{S}}G(\theta)\Psi(\theta)e^{-\kappa_iG(\theta)\|\lambda(\rho_i)\|}d\theta\to R_\epsilon.\tag{33}$$

If κ_c failed to be in $\mathfrak{K}_+^*$, then the sequence of continuous functions κ_iG would be negative on a subset of $\mathcal{S}$ of nonzero measure, for i large enough. This, together with the fact that $\|\lambda(\rho_i)\|\to\infty$, and the fact that the integral in (33) ought to be uniformly bounded, leads to a contradiction. If on the other hand, κ_c was an interior point of $\mathfrak{K}_+^*$ the sequence κ_iG would be strictly positive on $\mathcal{S}$, for i large enough. This, together with the growth of $\lambda(\rho_i)$ would indicate that the integral in (33) ought to tend to 0 as $i\to\infty$ instead of R_1. This again is a contradiction. Thus, $k_cG\ge 0$ on $\mathcal{S}$ except for a subset $\mathcal{S}_0$ of possibly zero measure where it vanishes.

In order to prove that $R_\epsilon\notin\text{int}(\mathfrak{K}(\mathfrak{G}))$ it suffices to show that the sequence R_{ρ_i}, which converges to R_ϵ from within $\text{int}(\mathfrak{K}(\mathfrak{G}))$, is such $\mathfrak{C}_{R_{\rho_i}}(\kappa_c)\to 0$ as $i\to\infty$. We readily observe that given any neighborhood $\mathcal{U}(\mathcal{S}_0)$ of $\mathcal{S}_0$,

$$\int_{\mathcal{S}/\mathcal{U}(\mathcal{S}_0)}e^{-\lambda(\rho_i)G(\theta)}d\theta=\int_{\mathcal{S}/\mathcal{U}(\mathcal{S}_0)}e^{-\|\lambda(\rho_i)\|\kappa_iG(\theta)}d\theta\to 0$$

because $\kappa_iG\to\kappa_cG$ uniformly on $\mathcal{S}$ and $\|\lambda(\rho_i)\|\to\infty$. On the other hand, $e^{-\lambda(\rho_i)G}$ are positive integrable functions, with a uniform bound on $\int_{\mathcal{S}}e^{-\lambda(\rho_i)G(\theta)}d\theta$ (due to (31) and the fact that R_ρ is bounded). But κ_cG vanishes on $\mathcal{S}_0$. Thus,

$$\int_{\mathcal{U}(\mathcal{S}_0)}\kappa_cG(\theta)\Psi(\theta)e^{-\|\lambda(\rho_i)\|\kappa_iG(\theta)}d\theta$$

can be made arbitrarily small with appropriate selection of $\mathcal{U}(\mathcal{S}_0)$. Hence, both terms in the representation

$$\begin{aligned}\mathfrak{C}_{R_{\rho_i}}(\kappa_c)\ &=\ \int_{\mathcal{S}/\mathcal{U}(\mathcal{S}_0)}\kappa_cG(\theta)\Psi(\theta)e^{-\|\lambda(\rho_i)\|\kappa_iG(\theta)}d\theta\\&\quad+\int_{\mathcal{U}(\mathcal{S}_0)}\kappa_cG(\theta)\Psi(\theta)e^{-\|\lambda(\rho_i)\|\kappa_iG(\theta)}d\theta\end{aligned}$$

can be made arbitrarily small for a suitable choice of $\mathcal{U}(\mathcal{S}_0)$ and sufficiently large values of i. This shows that $\mathfrak{C}_{R_\epsilon}$ is not strictly positive and, as a consequence, R_ϵ cannot be in the interior of the convex cone $\mathfrak{K}(\mathfrak{G})$. ■

To recap, we have shown that whenever R_1 admits a representation $\int_{\mathcal{S}}G(\theta)d\mu(\theta)$ with $d\mu$ a strictly positive measure, then one such measure can be found within the family of each theorem (with the dimensionality and periodicity restriction on G in the case of Theorem 2). This can be done by following the $\lambda(t)$-coefficients, which, as functions of the homotopy parameter t (or ρ), satisfy the differential equation given in the respective theorem.

A proof of uniqueness of such a representation is as follows: consider the mapping

$$\begin{aligned}h\ &:\ \mathcal{D}\to\text{int}(\mathfrak{K}(\mathfrak{G}))\\&\qquad\lambda\mapsto R=\int G(\theta)m(\theta,\lambda)d\theta\end{aligned}$$

where $\mathcal{D}$ refers to domains $\text{int}(\mathfrak{K}_+^*)$ and $\mathbb{R}^{n+1}$, respectively, for the two theorems. This is a C^1-mapping between two open convex subsets of $\mathbb{R}^{n+1}$, with a positive definite Jacobian throughout. Thus, if there exist two distinct vectors λ_1 and λ_2 in $\mathcal{D}$ which are mapped onto the same point $R_1 \in \text{int}(\mathfrak{K}(\mathfrak{G}))$, we may consider the path $\lambda_\rho := (1-\rho)\lambda_1 + \rho\lambda_2$ between them, for $\rho \in [0,1]$, and denote $R_\rho = h(\lambda_\rho)$. Since $\int_0^1 \frac{dR_\rho}{d\rho} d\rho = 0$ by the assumption that the two end points coincide, it follows that $(\int_0^1 (\frac{\partial h}{\partial \lambda}\Big|_{\lambda_\rho}) d\rho)(\lambda_2 - \lambda_1) = 0$ as well. But this is not possible since $\frac{\partial h}{\partial \lambda}$ is non-negative throughout, unless of course $\lambda_1 = \lambda_2$, which is the desired conclusion.

Remark 7: It is clear that the rational family of measures is essentially restricted to the 1-dimensional case of $\mathcal{S} = [0,1]$. It extends to the 2-dimensional setting only under the assumption that $G(\theta)$ is doubly periodic. A similar conclusion was drawn in [41, page 886]. The exponential family however has no such restriction. Alternative rational families, e.g., $1/(\langle \lambda, G\rangle)^3 d\theta$ etc., may turn out to be applicable for certain cases in the multidimensional setting. However, periodicity of $G(\theta)$ seems essential, otherwise, vanishing of $\langle \lambda, G\rangle$ may introduce only a single pole on the boundary of $\mathcal{S}$ for the density function, which may not be enough to prevent integrability in higher dimensional cases. □

Thus, the one-parameter imbedding leads to the differential equations (25) and (28), in the parameter ρ, which converge to the sought parameter λ_1 when integrated over the interval $[0,1]$ starting from λ_0. Replacing $R_1 - R_0$ by $\frac{1}{1-\rho}(R_1 - R_\rho)$ leads to a "feedback form" of the differential equation, where the vector field explicitely involves the difference between the moment vector at the "current position" and the "target value" R_1. The resulting differential equations are conveniently expressed in the new variable $t = -\log(1-\rho)$ and integrated over $[0,\infty)$. Thus, we obtain (5) and (8) given in Theorems 2 and 3. Both are of the form

$$\frac{d\lambda}{dt} = f(\lambda), \tag{34}$$

the only difference between them being in the precise form of the function $f(\lambda)$. It should also be emphasized that an ordinary ODE solver is all that is needed. A final point to be emphasized is that convergence is exponentially fast with time-constant of 1, since the trajectory in the λ-coordinates follows the trajectory of the linear differential equation

$$\frac{dR(t)}{dt} = R_1 - R(t), \tag{35}$$

via the correspondence

$$\lambda(t) \mapsto R = \int G(\theta) m(\lambda(t),\theta) d\theta \tag{36}$$

for $m(\cdot,\cdot)$ any of the two possibilities given (see also the schematic in Figure 7). We now complete the proof of the relevant claim in Theorems 2 and 3, which we restate in the following proposition.

Proposition 8: The function

$$V(\lambda) := \|R_1 - \int G(\theta) m(\lambda,\theta) d\theta\|^2$$

is a Lyapunov function for the differential equation (34). Moreover, along trajectories of (34) it holds that

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

Proof: Clearly, V is continuous, differentiable, with $V(\lambda) \geq 0$ for $\lambda \in \mathcal{D}$, ($\mathcal{D}$ being $\text{int}(\mathfrak{K}_+^*)$ or $\mathbb{R}^{n+1}$ depending on whether $m(\theta)d\theta$ is in $\mathcal{M}_{\text{rat}}$ or $\mathcal{M}_{\text{exp}}$, respectively) and only vanishes at a possible stationary point of (34). For either choice of $m(\lambda(t),\theta)$ equation (34) is simply:

$$\frac{d\lambda}{dt} = (\frac{\partial R}{\partial \lambda})^{-1}(R_1 - \int G(\theta) m(\lambda(t),\theta) d\theta)$$

with $\frac{\partial R}{\partial \lambda}$ being the Jacobian of (36). Hence,

$$\begin{aligned}
\frac{dV(\lambda(t))}{dt} &= \frac{\partial V(\lambda(t))}{\partial \lambda}\frac{d\lambda}{dt} \\
&= -2(R_1 - \int G(\theta) m(\lambda(t),\theta) d\theta)' \left(\frac{\partial R}{\partial \lambda}\right) \frac{d\lambda}{dt} \\
&= -2(R_1 - \int G(\theta) m(\lambda(t),\theta) d\theta)' \\
&\quad \times (R_1 - \int G(\theta) m(\lambda(t),\theta) d\theta) \\
&= -2V(\lambda(t)),
\end{aligned}$$

which completes the proof. ■

VI. Motivating examples (continued)

This section is meant to highlight the fact that constructing a moment generating function from a set of moments is *not any more complicated* when the set of kernel functions lacks any apparent "shift" structure. To this end we follow up on the examples presented earlier and present numerical results on certain representative cases. The two families of measures, in Theorems 2 and 3, give similar results. Hence, our selection as to which ones to display has been somewhat arbitrary. In all cases we took $\Psi \equiv 1$, although, a selection of an arbitrary Ψ does not in any way burden the computational steps.

A. Power spectrum of input given output measurements (cont.)

Theorems 2 and 3 can be adopted to the case where the integration kernels have support on the real line $\mathbb{R}$ or the half line $\mathbb{R}_+$, provided these kernels possess a limit at $\pm\infty$, cf. [39, Chapter V][4]. This is not an unreasonable assumption from a practical viewpoint. In fact, this is the case for the example introduced in Section II-A.

[4]Theorem 3 can also be adopted to higher dimensional cases with support in $\mathbb{R}^k$ or in $\mathbb{R}_+^k$, provided the kernels are similary well-behaved.

We begin with low-pass "sensors" having transfer functions $H_k(i\omega) = 1/(1 + i\omega/\tau_k)$, $k \in \{1, 2, 3, 4\}$, with $\tau_k \in \{1/2, 2/3, 1, 2\}$, respectively. We assume a real-valued stationary zero-mean stochastic process $\{u(t)\}$ as the input to all four of them. We further assume that the only available data are the variance r_0 of $u(t)$ and the variances r_k $(k = 1, 2, 3, 4)$ at the output of each sensor. Hence, if $\mu(\omega)$ denotes the power distribution function of $\{u(t)\}$, these variances provide the following moment contraints:

$$r_k = \int_{-\infty}^{\infty} g_k(\omega) d\mu(\omega) \text{ for } k = 0, 1, \ldots, 4, \qquad (37)$$

for $g_0(\omega) = 1$ and

$$g_k(\omega) = \frac{1}{\omega^2/\tau_k^2 + 1}$$

with $\tau_k = 1/2, 2/3, 1, 2$ for $k = 1, 2, 3, 4$, respectively.

In order to compare a "reconstructed" power spectrum with an "authentic" one we take

$$d\mu_o(\omega) = \frac{1}{\omega^4 - 1.8\omega^2 + 1} d\omega$$

as the power distribution of $\{u(t)\}$. The spectral density function $m_o(\omega) := \dot{\mu}_o(\omega) = 1/(\omega^4 - 1.8\omega^2 + 1)$ is drawn in Figure 8 with a continuous line.

We readily compute the variances $r_0, \ldots, r_4$ and form the vector of moments

$$\begin{aligned} R_1 &= \begin{bmatrix} r_0 & r_1 & r_2 & r_3 & r_4 \end{bmatrix}' \\ &= \begin{bmatrix} 7.0250 & 4.1545 & 2.2579 & 5.8332 & 2.9938 \end{bmatrix}'. \end{aligned}$$

The next step is to seek a spectral density function that agrees with the moments in R_1.

The transformation

$$\omega \mapsto \theta = \frac{\omega}{1 + \omega}$$

takes the half axis $[0, \infty]$ onto $[0, 1]$ and brings us to a familiar setting. Its inverse is given by

$$\theta \mapsto \omega = \frac{\theta}{1 - \theta},$$

while the differentials relate via $d\omega = d\theta/(1 - \theta)^2$ and $d\theta = d\omega/(1 + \omega)^2$. For a general μ in (37), the moment constraints become

$$r_k = \int_0^1 \underbrace{2g_k(\frac{\theta}{1-\theta})}_{\hat{g}_k(\theta)} \underbrace{m(\frac{\theta}{1-\theta})\frac{1}{(1-\theta)^2}}_{\hat{m}(\theta)} d\theta$$

for $k \in \{0, 1, 2, 3, 4\}$. The framework of Theorems 2 and 3 now applies. The integration kernel functions are continuous on the closed interval $[0, 1]$. Moreover, $m(\omega)d\omega$ is a finite measure on $[0, \infty]$ provided $\hat{m}(\theta)d\theta$ is a finite measure on $[0, 1]$, and vice versa. Hence, a finite measure $\hat{m}(\theta)d\theta$ can be identified to match the given moments in either the rational or the exponential form (as per Section III) . Then,

$$m(\omega)d\omega = \frac{1}{(1+\omega)^2}\hat{m}(\frac{\omega}{1+\omega})d\omega$$

serves as an appropriate spectral measure for $\{u(t)\}$ which is consistent with the given moments.

We specialize to the context of Theorem 2 and integrate (5) so as to obtain a rational spectral density with support on $[0, 1]$ in the form

$$\hat{m}(\theta) = 1/(\sum_{k=0}^{4} \lambda_k \hat{g}_k(\theta))$$

which agrees with R_1. Then a spectral density for the original problem, as a function of ω, is obtained in the form

$$m(\omega) = \frac{1}{(\omega+1)^2}\hat{m}(\frac{\omega}{1+\omega}).$$

For the particular values selected earlier, the resulting $m(\omega)$ is drawn in Figure 8 with a dashed line, for comparison. Figure 9 shows the error $V(\lambda(t)) = \|R_1 - R_t\|^2$ in matching the known variances as a function of the integration variable t of Theorem 2–converging to zero with exponent -2. The lower subplot of Figure 9 shows how the entropy integral $\int_0^\infty \log(m_t(\omega))d\omega$ varies with t. In this example it is monotonically increasing, which is not a general property.

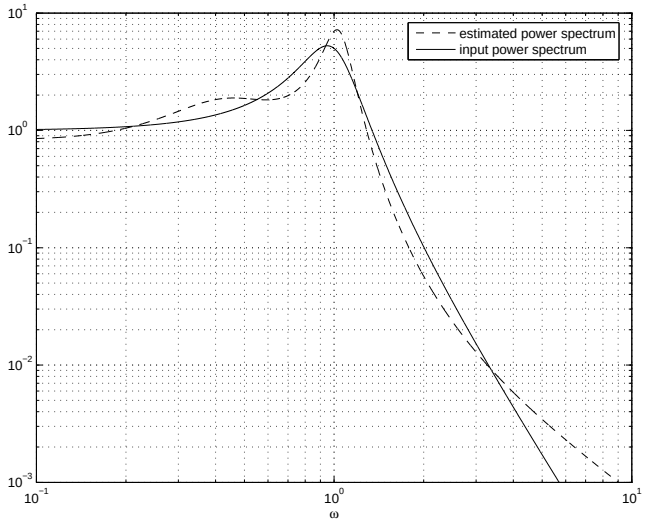

Fig. 8. Spectral density based on covariance statistics

B. Non-uniformly spaced sensor arrays - planar far field excitation (cont.)

We once again consider the linear array shown in Figure 1 and discussed in Section II-B. This consists of three non-equispaced sensors. We generate a vector of moments

$$\begin{aligned} R_1 &= \begin{bmatrix} r_0 & r_1 & r_\pi & r_{\pi+1} \end{bmatrix}' \\ &= \begin{bmatrix} 0.6210 & 0.5779 & 0.2776 & 0.1102 \end{bmatrix}' \end{aligned}$$

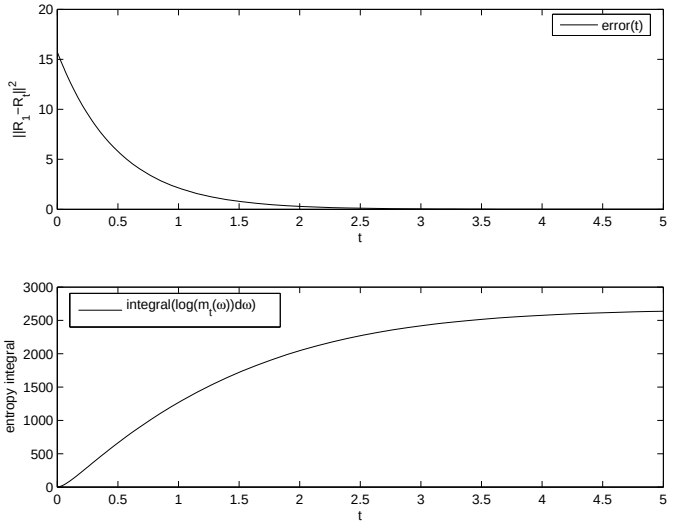

Fig. 9. Error $\|R_1 - R_t\|$ and $\int_0^\infty \log(m_t(\omega))d\omega$ vs. t

via

$$r_k \quad = \quad \int_0^1 \cos(k\theta) m_{\rm ref}(\theta) d\theta,$$

for $k \in \{0, 1, \pi, \pi+1\}$ and a particular "reference" spectral density $m_{\rm ref}$ chosen as piece-wise constant on $\mathcal{S} = [0, 1]$. Figure 10 shows $m_{\rm ref}$ with a dashed line as well as three other spectral density functions that are consistent with R_1. In particular, Figure 10 shows densities corresponding to the families $\mathcal{M}_{\rm exp}$ and $\mathcal{M}_{\rm rat}$ which have been obtained as indicated in Theorems 2 and 3, respectively, with $\Psi \equiv 1$. It can be seen that these two spectral densities differ very little from one another (the first marked with a continuous line and the second with a dotted line).

In the same figure we superimpose one additional spectral density marked as dash-dotted $(- \cdot -)$. This is listed as an element of constant + $\mathcal{M}_{\rm exp}$ (meaning that $m(\theta)d\theta$ = constant$d\theta$+ and element of the exponential family). This last spectral density is obtained by mimicking the Pisarenko-Carathèodory construction which underlies high resolution methods (see [55], [25], [26]). Briefly, if we postulate that the moment generating measure $d\mu$ has a "white noise" component $p_0 d\theta$, then it is possible to determine the maximal power density that such a component can account for in explaining R_1. To this end, we compute a vector of moments corresponding to a uniform measure $d\theta$. This turns out to be

$$\begin{aligned} R_{1,\rm white} &= \int_0^1 \left[\begin{array}{cccc} 1 & \cos(\theta) & \cos(\pi\theta) & \cos((\pi+1)\theta) \end{array}\right]' d\theta \\ &= \left[\begin{array}{cccc} 1 & 0.8415 & 0 & -0.2032 \end{array}\right]'. \end{aligned}$$

Then,

$$p_0 = \text{argmax}\{p \ : \ R_1 - pR_{1,\rm white} \in \mathfrak{K}\}.$$

For each value of p that $R_1 - pR_{1,\rm white} \in \text{int}(\mathfrak{K})$ we can apply either Theorem 2 or 3, and obtain a spectral measure $d\mu$ for which

$$R_1 = \int_0^1 G(\theta)(pd\theta + d\mu(\theta)).$$

As we increase p and $R_1 - pR_{1,\rm white}$ approaches the boundary of $\mathfrak{K}$, the measure $d\mu$ tends to become singular with singularities at roots of $\langle \lambda, G\rangle$ for a suitable λ on the boundary of $\mathfrak{K}_+^*$ (cf. Remark 4). This is indeed the case when $p \sim 1.005$ (the value used for the figure is 1.004).

It may be instructive to draw the connection with case where G is an ordinary Fourier vector corresponding to an equispaced linear array uniform uniform spacial sampling, i.e., when $g_k(\theta) = \cos(k\theta)$. Then p_0 is the lowest eigenvalue of a Toeplitz matrix formed out of the moments.

The same rationale can be used to determine the maximal power density for a component of known "color" that is consistent with the given moments, cf. [26]. Thus, using such ideas and further exploiting the parameter Ψ in the two theorems (Theorems 2 and 3) a variety of spectral density functions can be generated—all consistent with the moments—that may incorporate prior information about the moment generating density/measure.

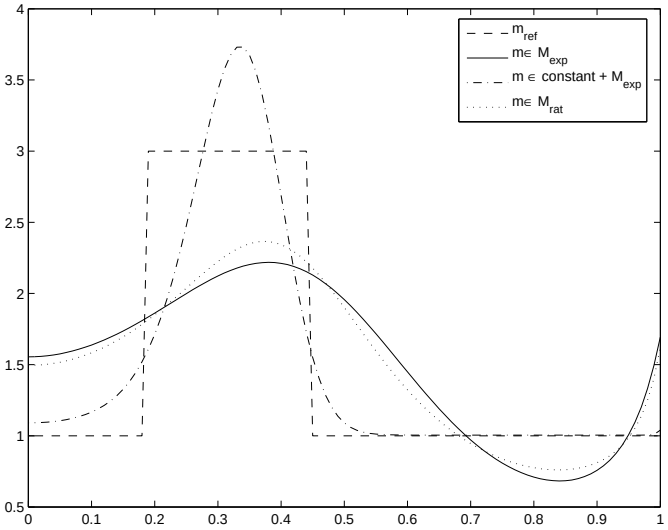

Fig. 10. $d\mu \in \mathcal{M}_{\rm exp}$

We conclude by demonstrating the effect of additional moments. The same $m_{\rm ref}(\theta)d\theta$ is used to generate moments for an array extended with two additional elements, say E_4 and E_5, in line with the rest and at distances 9 and 19 wavelenghts from E_1. A schematic is shown in Figure 11. Utilizing the first k elements, with $k \in \{3, 4, 5\}$, and correlating the respective readings at each sensor location, we have

$$G_k(\theta) = \left[1 \quad \cos(d_1\theta) \quad \ldots \quad \cos(d_\ell\theta) \quad \ldots \quad \cos(d_{\frac{k(k+1)}{2}}\theta)\right]'$$

as the "array manifold" with d_ℓ taking values in

$$\begin{aligned} &\{0, 1, \pi+1, \pi\}, \\ &\{0, 1, \pi+1, 9, \pi, 8, 8-\pi\} \text{ and} \\ &\{0, 1, \pi+1, 9, 19, \pi, 8, 18, 8-\pi, 18-\pi, 10\}, \end{aligned}$$

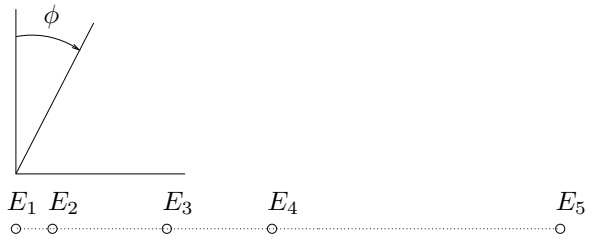

Fig. 11. Geometry of sensor array

respectively, in the three cases. The values correspond to distances between every pair of elements of the sensor array in each case. Figure 12 shows for comparison spectral density functions $m_3(\theta)$, $m_4(\theta)$ and $m_5(\theta)$, constructed according to Theorem 2, i.e., with the corresponding measures $m_k(\theta)d\theta \in \mathcal{M}_{\rm rat}$ $(k = 3, 4, 5)$. Each shares with $m_{\rm ref}(\theta)$ $\frac{k(k+1)}{2}+1$ moments, in each case. It is noted that, as expected, matching of m_k and $m_{\rm ref}$ improves as k increases.

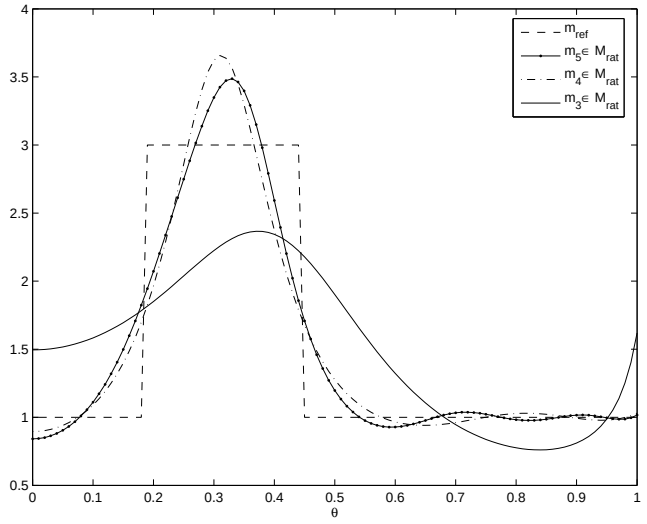

Fig. 12. Non-uniformly spaced moments: using $\mathcal{M}_{\rm rat}$

C. Sensor arrays – multidimensional case

Theorem 3 can be used to generate all multidimensitonal distributions that match given moments, in complete generality. However, the example we have selected to demonstrate its applicability is rather basic. The support set $\mathcal{S}$ is two dimensional, the convolution kernels are sinusoidal functions (though not periodic in $\mathcal{S}$), and we are content with producing the one distribution which corresponds to selection of $\Psi = 1$.

We begin with a two dimensional density function

$$m_{\rm ref}(\theta,\phi) := 10 + 10e^{-7\left((\theta-\theta_0)^2+(\phi-\phi_0)^2\right)}$$

with $(\theta,\phi) \in \mathcal{S} := [0,\pi]\times[0,\pi]$, and centered at $(\theta_0,\phi_0) = (1.5, 1.5)$. The integration kernels are chosen as

$$g_{k,\ell}(\theta,\phi) = \cos(k\theta + \ell\phi)$$

for $k, \ell \in \{0, 1, 2\}$, and a corresponding matrix of moments is computed as

$$R_1 = \left[\int_{\mathcal{S}} g_{k,\ell}(\theta,\phi)m_{\rm ref}(\theta,\phi)d\theta d\phi\right]_{k,\ell=0}^{k,\ell=2} = \begin{bmatrix} 33.0129 & 0.3140 & -1.1417 \\ 0.3140 & -14.0469 & -0.2502 \\ -1.1417 & -0.2502 & 1.0310 \end{bmatrix}$$

This formulation encompasses spatial correlation data of a two-dimensional array (see [35, page 50]), neglecting temporal correlations, where the support of the impinging (narrow bandwidth stationary, etc.) wave is parametrized by the two angular parameters θ, ϕ. The particular $m_{\rm ref}$ may represent a dominant component concentrated around $(\theta_0, \phi_0) = (1.5, 1.5)$ on a uniform background.

We continue on as usual with

$$\lambda_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We integrate (8) with Ψ equal to 1 and

$$\langle \lambda, G(\theta,\phi)\rangle := \text{trace}(\lambda' G(\theta,\phi)).$$

Since R_1 consists of moments of a positive distribution, the differential equation converges as claimed in the theorem, and the limiting value of $\lambda(t)$ turns out to be

$$\lambda = \begin{bmatrix} -2.3720 & 0.0005 & 0.0909 \\ 0.0005 & -0.0715 & 0.0117 \\ 0.0909 & 0.0117 & -0.0916 \end{bmatrix}.$$

As claimed in the theorem, the corresponding density

$$e^{-\langle\lambda,G\rangle}$$

agrees with the moments in R_1. The two density functions $m_{\rm ref}$ and $e^{-\langle\lambda,G\rangle}$ are shown in Figures 13 and 14, respectively, for comparison.

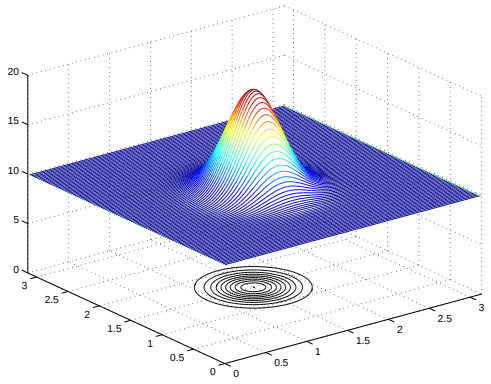

Fig. 13. $m_{\rm ref}(\theta,\phi)$ on $[0,\pi]\times[0,\pi]$

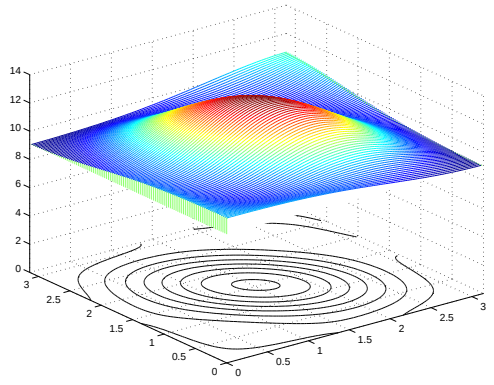

Fig. 14. $e^{-\langle \lambda, G(\theta,\phi)\rangle}$ on $[0,\pi]\times[0,\pi]$

VII. Synopsis

This paper develops a theory for the general scalar[5] moment problem. The formalism is sufficiently general to encompass problems in system identification, sensor arrays with arbitrary geometry and dynamics, and in nonuniform multi-dimensional sampling. It begins with a known finite set of moments with respect to known convolution kernels. The goal is to determine whether the moments are consistent with the hypothesis of a positive moment-generating density, and if so, to characterize all densities which are consistent with the moments.

Tools and techniques of the classical theory of the moment problem and of analytic interpolation are not suitable for the generality sought in the paper—where the convolution kernels have no discernible structure and the support of the density function can be multidimensional. Hence, our approach is quite distinct from the classical theory. Yet, it answers the typical questions of existence and parametrization of solutions just as effectively and in great generality.

The techniques we have used complement those in earlier works based on duality and convexity theory [22], [44], [36], [13]. They are also quite distinct from homotopy methods employed in [11], [20], [48], [9] in a specialized context.

VIII. Acknowledgments

The present work has benefitted from discussions with Laurent Baratchart during a leave at the Mittag-Leffler institute in Sweden, in the Spring of 2003, and from earlier collaborative work of the author with Chirs Byrnes and Anders Lindquist.

References

[1] N. I. Akhiezer, **The classical moment problem,** Hafner Publishing, Translation 1965, Oliver and Boyd.

[5] A follow-up paper [30] explains how the theory extends to matrix-valued measures and their moments $R = \int_{\mathcal{S}} (G_{\text{left}} d\mu G_{\text{right}})$ with respect to matrix-valued kernels G_{left} and G_{right}. It is shown that a "rational" family of matricial density functions of the form $\Psi^{1/2}\left((G_{\text{right}}\lambda G_{\text{left}})_{\text{Hermitian}}\right)^{-1}\Psi^{1/2}$ allows all conclusions of Theorem 2 to go through in the multivariable setting, and that an "exponential" family $\Psi^{1/2}\exp\left(-(G_{\text{right}}\lambda G_{\text{left}})_{\text{Hermitian}}\right)\Psi^{1/2}$ allows extending the conclusions of Theorem 3. These are in fact minimizers of a matricial version of relative entropy.

[2] N. I. Akhiezer and M. G. Krein, **Some Questions in the Theory of Moments,** American Mathematical Society, Providence, RI, 265 pages, 1962.

[3] D. Z. Arov, "The generalized bitangent Carathéodory-Nevanlinna-Pick problem, and (j, J_0)-inner matrix-valued functions," *Russian Acad. Sci. Izv. Math.*, **42(1)**: 1-26, 1994.

[4] D. Z. Arov and H. Dym, "On three Krein extension problems and some generalizations," *Integral Equations and Operator Theory*, **31**: 1-91, 1998.

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

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

[7] J.M. Borwein and A.S. Lewis, "Duality relationships for entropy-like minimization problems," *SIAM J. Control Optim.* **29**: 325-338, 1991.

[8] J.M. Borwein and A.S. Lewis, "Partially-finite programming in L_1 and the existence of maximum entropy estimates," *SIAM J. Optim.* **2**: 248-267, 1993.

[9] A. Blomqvist, G. Fanizza, and R. Nagamune, "Computation of bounded degree nevanlinna-pick interpolants by solving nonlinear equations," Proceedings of the 42nd IEEE Conference on Decision and Control, vol. 5: 4511 - 4516,Dec. 2003.

[10] 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.

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

[12] 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.

[13] C. I. Byrnes and A. Lindquist, "Interior point solutions of variational problems and global inverse function theorems," KTH/OPT Report ISSN 0348-405X.

[14] 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**, pages 3-21, Birhäuser, 2003.

[15] 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.

[16] R.E. Curto and L.A. Fialkow, "The truncated complex K-moment problem," *Trans. of the American Math. Society*, **352(6)**: 2825-2855, 2000.

[17] D. Dacunha-Castelle et F. Gamboa, "Maximisation de l'entropie sous contraintes non linèaires," Annales de l'Inst. H. Poincarre, **4**: 567-596, 1990.

[18] B.W. Dickinson, "Two dimensional Markov spectrum estimates need not exist," *IEEE Trans. on Information Theory*, **IT-26**: 120-121, 1980.

[19] H. Dym, **J-contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation**, CBMS vol. 71, American Math. Soc., 1989.

[20] P. Enqvist, "A homotopy approach to rational covariance extension with degree constraint," *Intern. J. Applied Mathematics and Computer Science* **11(5)** (2001), 1173–1201.

[21] M. Frontini and A. Tagliani, "Maximum entropy in the generalized moment problem," *Journal of Math. Physics*, **39(12)**: 6706-6714, 1998.

[22] 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.

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

[24] T.T. Georgiou, "Realization of power spectra from partial covariance sequences," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, AASP-35(4): 438-449, April 1987.

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

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

[27] 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.

[28] T.T. Georgiou, "Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization," *IEEE Trans. on Automatic Control*, November 2002, **47(11):** 1811-1823.

[29] T.T. Georgiou, "A differential equation for the LMI feasibility problem," preprint, July 2004.

[30] T.T. Georgiou, "The problem of testing positivity of covariance samples and of realizing consistent power spectra," preprint, June 2004; "Relative Entropy and the multi-variable multidimensional Moment Problem" (revised title), December 2004.

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

[32] T.T. Georgiou, P.J. Olver, and A. Tannenbaum, "Maximal entropy for reconstruction of back projection images," "IMA Volumes in Mathematics and its Applications," *Volume 133: Mathematical methods in computer vision* Springer-Verlag, New York, 2002.

[33] U. Grenander and G. Szegö, **Toeplitz Forms and their Applications**, Chelsea, 1958.

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

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

[36] M. Junk, "Maximum entropy for reduced moment problems," *Mathematical Models and Methods in Applied Sciences*, **10(7)**: 1001-1025, 2000.

[37] S. Karlin and W. Studden, **Tchebycheff systems: With applications in analysis and statistics. Pure and Applied Mathematics**, Vol. XV Interscience Publishers John Wiley and Sons, New York-London-Sydney 1966.

[38] R.B. Kellogg, T.Y. Li, and J. Yorke, "A constructive proof of the Brouwer fixed-point theorem and computational results," *SIAM J. Numer. Anal.*, **13(4)**: 739-752, September 1976.

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

[40] H.J. Landau, "Maximum entropy and the moment problem," *Bulletin (New Series) of the American Math. Society*, **16(1)**: 47-77, 1987.

[41] 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.

[42] R.D. Levine and M. Tribus (editors), **The Maximum Entropy Formalism**, MIT Press, Cambridge, 1979.

[43] 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.

[44] A.S. Lewis, "Consistency of moment systems," *Can. J. Math.*, **47(5)**: 995-1006, 1995.

[45] T.L. Marzetta, "Two-dimensional linear prediction: autocorrelation arrays, minimum-phase predition error filters, and reflection coefficient arrays," *IEEE Trans. on Acoustics Speech and Signal Proc.*, **ASSP-28(6)**: 725-733, 1980.

[46] G.H. Meyer, "On solving nonlinear equations with a one-parameter operator imbedding," *SIAM J. Numer. Anal.*, **5(4)**: 739-752, December 1968.

[47] 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," *Proceedings of the IEEE*, **75(7)**: 892-907, 1987.

[48] R. Nagamune, "A Robust Solver Using a Continuation Method for Nevanlinna-Pick Interpolation with Degree Constraint," *IEEE Transactions on Automatic Control*, **48(1)**: 113-117, January 2003.

[49] M. Putinar, "Extremal solutions of the two dimensional L-problem of moments," Journal of Functional Analysis, **136**: 331-364, 1996.

[50] W. Rudin, "The extension problem for positive-definite functions," *Illinois J. Math.*, **7**:532-539, 1963.

[51] W. Rudin, **Functional Analysis**, 2nd edition, McGraw Hill, 1991.

[52] D. Sarason, Generalized Interpolation in $\mathcal{H}^\infty$, *Transactions of the American Math. Society*, **127:** 179-203, 1967.

[53] J.A. Shohat and J.D. Tamarkin, **The problem of moments,** New York city, American mathematical society, 1943, 140 pages, 1943.

[54] K.V. Stavropoulos and A. Manikas, "A game theory H_∞ approach to robust direction finding," Proceedings of ICSP2000, pages 2128-2131, 2000.

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

[56] T. Stieljes, "Recherches sur les fractions continues," Ann. Fac. Sci. Univ. Toulouse, **8**: J1-J122, 1894; **9**: A1-A47, 1895; reprinted in *Oeuvres compleètes,* tomes 1, 2, Noordhoff, Groningen, 1914, 1918.

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

[58] J.W. Woods, "Two-dimensional Markov spectral estimation," *IEEE Trans. on Information Theory,* **IT-22**: 552-559, 1976.