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

Signal analysis, moment problems & uncertainty measures

Johan Karlsson and Tryphon T. Georgiou

***Abstract*— Modern spectral estimation techniques often rely on second order statistics of a time-series to determine a power spectrum consistent with data. Such statistics provide moment constraints on the power spectrum. In this paper we study possible distance functions between spectra which permit a reasonable quantitative description of the uncertainty in moment problems. Typically, there is an infinite family of spectra consistent with given moments. A distance function between power spectra should permit estimating the diameter of the uncertainty family, a diameter which shrinks as new data accumulates. Abstract properties of such distance functions are discussed and certain specific options are put forth. These distance functions permit alternative descriptions of uncertainty in moment problems. While the paper focuses on the role of such measures in signal analysis, moment problems are ubiquitous in science and engineering, and the conclusions drawn herein are relevant over a wider spectrum of problems.**

I. INTRODUCTION

The moment problem in its most basic formulation amounts to determining a non-negative distribution $d\mu$ consistent with a given set of moments

$$\int_{\theta\in\mathcal{S}} g_k(\theta)d\mu(\theta),\ \text{for } k=0,1,2,\ldots,n,$$

where $d\mu$ and the integration kernels g_k are defined on a support set $\mathcal{S}$. The classical theory [16], [17], [1] focused on $\mathcal{S}$ being 1-dimensional and on the integration kernels forming a Tchebyshev system (see [16]). It is worth noting that analytic interpolation problems of the Nevanlinna-Pick type can be cast as such [16]. Perhaps the most commonly encountered version of such a problem is the so-called *trigonometric moment problem* where $g_k(\theta)=e^{-jk\theta}$, $\theta\in(-\pi,\pi]$ and $d\mu(\theta)$ a bounded non-negative measure. It is this latter problem that will be our focus in this paper.

Motivation for studying the trigonometric moment problem comes from the theory of stochastic processes. Indeed, if $\{y_t\ :\ t\in\mathbb{Z}\}$ is a discrete time, zero mean, second order stationary stochastic process, the auto-covariances (i.e., autocorrelation samples)

$$c_k := E(y_t\bar{y}_{t-k}),\ \text{for } k=0,\pm1,\pm2,\ldots,\pm n$$

provide moment constraints on the power spectrum $d\mu$ of the process:

$$c_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-ik\theta}d\mu(\theta)\ \text{for } k=0,\pm1,\pm2\ldots,\pm n. \quad (1)$$

It is well known [12], [13] that moments of a non-negative distribution (at least for classical problems) are characterized by the non-negativity of a suitable quadratic form—in our case, the form specified by the Toeplitz matrix

$$T_n=\begin{bmatrix} c_0 & c_{-1} & \cdots & c_{-n}\\ c_1 & c_0 & \cdots & c_{-n+1}\\ \vdots & \vdots & \ddots & \vdots\\ c_n & c_{n-1} & \cdots & c_0\end{bmatrix}.$$

When $T_n\geq 0$ and singular, there is a unique $d\mu$ consistent with (1) and the requirement that $d\mu\geq 0$. In fact, this $d\mu$ is singular with respect to the Lebesgue measure and consists of a finite number ($\leq n$) of "spectral lines." In general however, when $T_n>0$, the family of consistent distributions

$$\mathcal{F}_{\mathbf{c}_{0:n}}=\{d\mu :\ d\mu\geq 0,\ \text{and (1) holds}\}$$

is an infinite one. Here, and throughout, $\mathbf{c}_{0:n} := (c_0, c_1,\ldots, c_n)$ denotes the vector of the first $(n+1)$ moments, $\mathbf{c} := (c_0, c_1,\ldots)$ denotes the infinite sequence, while the fact that $d\mu$ is a real measure dictates that $c_k=\bar{c}_{-k}$ for $k=0,1,\ldots$. The sequence $\mathbf{c}$ is said to be *positive* if $T_n>0$ for all n. Similarly $\mathbf{c}_{0:n}$ is said to be *positive* if $T_n>0$. Accordingly, the term *non-negative* is used when the relevant Toeplitz matrices are non-negative definite.

Theory and practice of spectral estimation (see [13], [18]) revolve around specific choices within $\mathcal{F}_{\mathbf{c}_{0:n}}$ which then form the basis of particular spectral estimation algorithms (see [4] for a concise exposition). *In the present paper we view $\mathcal{F}_{\mathbf{c}_{0:n}}$ as an "uncertainty set"*. We seek ways to quantify "modeling uncertainty" and "variability of spectra" consistent with (1). More specifically, we seek suitable distance functions

$$\delta(d\mu_1,d\mu_2)\geq 0$$

with *certain natural properties* that allow defining a "diameter"

$$\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}}) := \sup\{\delta(d\mu_1,d\mu_2)\ :\ d\mu_1,d\mu_2\in\mathcal{F}_{\mathbf{c}_{0:n}}\}$$

of $\mathcal{F}_{\mathbf{c}_{0:n}}$.

The following properties are sought for any such distance function $\delta(\cdot,\cdot)$. First, that the corresponding induced radius $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ decreases to zero whenever $\mathcal{F}_{\mathbf{c}_{0:n}}$ tends towards a singleton. There are two cases in particular when this happens. This is the case when the length of the vector

Supported by Vetenskapsrådet, the National Science Foundation and the Air Force Office of Scientific Research

Johan Karlsson is with the Department of Mathematics, Division of Optimization and Systems Theory, Royal Institute of Technology, 100 44 Stockholm, Sweden, johan.karlsson@math.kth.se

Tryphon T. Georgiou is with the Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA, tryphon@ece.umn.edu

0-7803-9568-9/05/$20.00

of moments $\mathbf{c}_{0:n}$ increases without bound, i.e., when c_k for $k = n+1, n+2, \ldots$ become successively known. This is due to the fact that the trigonometric problem is *determinate* [1] in that $\mathbf{c}$ specifies uniquely a measure $d\mu$ for which (1) holds for all n, i.e.,

$$\cap_{n=0}^{\infty} \mathcal{F}_{\mathbf{c}_{0:n}} = \{d\mu\}.$$

A second case is when the values for the first $n+1$ moments is perturbed so that the Toeplitz matrix T_n tends to become singular. In both those cases $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ needs to reflect that the size decreases to 0. It turns out that these first two properties are closely related to a certain type of continuity of $\delta(\cdot,\cdot)$ (weak* continuity). An additional issue we consider is that of computability of $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$. It turns out that computability is greatly facilitated by convexity of $\delta(\cdot,\cdot)$.

Section II reviews certain mathematical concepts on measures and harmonic functions as they pertain to representation of power spectra. Section III expands on the connection between distance functions and the induced diameter of the uncertainty set. Section IV gives a characterization of the essential boundary of uncertainty sets, while Sections V and VI expand on possible alternatives for distance functions with desirable properties. Technical arguments and background information are provided in an appendix.

II. Power spectra & harmonic analysis

The power spectrum of a discrete-time stationary process can be thought of as a bounded non-negative measure on the unit circle. The derivative (of its absolutely continuous part) is often referred to as spectral density while the singular part may contain jumps (spectral lines) associated with the presence of sinusoidal components. Non-negative measures are naturally associated with analytic and harmonic functions—a connection which has profitably been studied in classical circuit theory in the context of passive circuits. In fact, power spectra relate, in a very precise sense, to boundary limits of the (harmonic) real parts of so-called "positive-real functions." This brief section reviews relevant facts and notation.

Weak convergence*

Bounded measures on the boundary of the unit disc $\mathbb{D} := \{z \;:\; |z| < 1\}$ can be thought of as functionals on $C(\mathbb{T})$, the class of continuous functions defined on $\mathbb{T} := (-\pi, \pi]$. Indeed, if $C(\mathbb{T})^*$ denotes the set of such bounded linear functionals $\Lambda : C(\mathbb{T}) \to \mathbb{R}$, the Riesz representation theorem asserts the existence of bounded measure $d\mu$ such that

$$\Lambda(f) = \int_{\mathbb{T}} f(t)d\mu(t)$$

for all $f \in C(\mathbb{T})$. Then, a natural topology is induced by the notion of *weak* convergence*, where a sequence $d\mu_n$ converges to $d\mu$ if $\int f d\mu_n \to \int f d\mu$ for every $f \in C(\mathbb{T})$.

Analytic and harmonic representations of power spectra

Herglotz' theorem states that if $d\mu$ is a bounded non-negative measure on $\mathbb{T}$, then

$$H[d\mu](z) = \frac{1}{2\pi}\int_{\mathbb{T}} \frac{e^{it}+z}{e^{it}-z} d\mu(t)$$

defines a function which is analytic in $\mathbb{D}$ and has positive real part. Conversely, any such function can be represented (modulo an imaginary constant) by the above formula for a suitable non-negative measure. The class of analytic functions in $\mathbb{D}$ with non-negative real part is usually denoted by $\mathcal{C}$ (after Carathèodory) and this is the notation we follow as well.

The Poisson integral of a non-negative measure $d\mu$

$$P[d\mu](z) := \frac{1}{2\pi}\int_{\mathbb{T}} P_r(\theta - t)d\mu(t), \quad z = re^{i\theta},$$

where $P_r(\theta) = \frac{1-r^2}{|1-re^{i\theta}|^2}$ is the Poisson kernel, is a harmonic function which is non-negative in $\mathbb{D}$ and equal to the real part of $H[d\mu](z)$. The measure $d\mu$ is uniquely determined via the limit of $P[d\mu](re^{i\theta})d\theta \to d\mu$ as $r \to 1$ in the weak* topology.

III. Limit properties of the diameter ρ_δ

The point of this section is to show that weak* continuity of the distance function δ gives the desired limit properties for the radius $\rho_\delta(\mathcal{F}_c)$ of the uncertainty set.

We begin by highlighting the fact that weak* convergence of measures relates to convergence of their Poisson integrals.

Proposition 1: Let $\{d\mu_k\}_{k=1}^{\infty}$ be a sequence of uniformly bounded signed measures on $\mathbb{T}$, let $d\mu$ be a bounded measure on $\mathbb{T}$, and let $u(z) = P[d\mu](z)$, $u_k(z) = P[d\mu_k](z)$ be their corresponding Poisson integrals. The following statements are equivalent:

1) $d\mu_k \to d\mu$ weak*,
2) $u_k(z) \to u(z)$ pointwise $\forall z \in \mathbb{D}$,
3) $u_k(z) \to u(z)$ in $L_1(\mathbb{D})$,
4) $u_k(z) \to u(z)$ uniformly on compact subsets of $\mathbb{D}$.

Proof: The equivalence of the conditions follows from standard arguments of complex analysis (see [14]). ■

Formally, in this paper, a *distance function* is simply a mapping from a pair of non-negative measures to the non-negative reals , i.e., $(d\mu_1, d\mu_2) \mapsto \delta(d\mu_1, d\mu_2) \geq 0$, such that the non-negative measures with metric δ is a metric space. The next proposition points out that weak*-continuity of the distance function ensures that the induced diameter of $\mathcal{F}_{\mathbf{c}_{0:n}}$ goes to zero as $n \to \infty$.

Proposition 2: If $\delta(\cdot,\cdot)$ is a weak* continuous distance function, and $\mathbf{c}$ a non-negative sequence, then $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}}) \to 0$ as $n \to \infty$.

Proof: Let $d\mu$ be the unique measure specified by $\mathbf{c}$, and let $d\mu_n \in \mathcal{F}_{\mathbf{c}_{0:n}}$ denote a sequence of measures, for $n = 1, 2, \ldots$. Clearly $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$ for all n. From [10, §1.16] it follows that

$$|P[d\mu](z) - P[d\mu_n](z)| \leq \frac{c_0\sqrt{8}|z|^n}{(1-|z|)^{\frac{3}{2}}}, \forall n,$$

and hence from Proposition 1, that $d\mu$ is the weak* limit of $d\mu_n$ as $n \to \infty$. Weak* continuity of δ now ensures that the supremum of $\delta(d\mu_n, d\mu)$ over $d\mu_n \in \mathcal{F}_{\mathbf{c}_{0:n}}$ goes to zero with n, and so does $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$. ■

The following proposition explains what happens when the values of the entries of $\mathbf{c}_{0:n}$, as a vector of constant length n, vary so that T_n tends to become singular.

Proposition 3: Let $\hat{\mathbf{c}}_{0:n}$ be a non-negative sequence corresponding to a singular Toeplitz matrix (i.e., $\hat{\mathbf{c}}_{0:n}$ in non-negative but not positive), and let $\mathbf{c}_{0:n}(\ell)$ ($\ell = 1, 2, \ldots$) denote a sequence of nonnegative $(n+1)$-vectors of moments tending to $\hat{\mathbf{c}}_{0:n}$. If $\delta(\cdot, \cdot)$ is a weak* continuous distance function, then $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}(\ell)}) \to 0$ as $\ell \to \infty$.

Proof: See appendix. ■

It should be noted that if $\mathbf{c}_{0:n}$ is positive, then $\mathcal{F}_{\mathbf{c}_{0:n}}$ contains infinitely many measures and among them at least two singular measures with non-overlapping support, i.e., $\mathrm{supp}(d\mu_1) \cap \mathrm{supp}(d\mu_2) = \emptyset$ and the total variation of their difference is $2c_0$. Therefore, the statements of Proposition 2 and Proposition 3 would fail if the total variation is used to define distances in the cone of non-negative measures.

IV. Extreme points of the uncertainty set

Computation of the diameter $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ of the uncertainty set amounts to solving the optimization problem $\sup\{\delta(d\mu_1, d\mu_2) \, : \, d\mu_1, d\mu_2 \in \mathcal{F}_{\mathbf{c}_{0:n}}\}$. At the outset this appears infinite dimensional, since an infinite set of parameters are needed to characterize a typical element $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$. However, it turns out, that weak* continuity and convexity of δ reduce the problem to a finite dimensional one that can be solved with standard methods.

The uncertainty set $\mathcal{F}_{\mathbf{c}_{0:n}}$ is the intersection of the positive cone (non-negative measures) with a closed subspace (moment constraints), hence it is convex and closed. The norm of its elements is bounded by c_0—this is the total variation since these are considered as functionals on $C(\mathbb{T})$. Hence $\mathcal{F}_{\mathbf{c}_{0:n}}$ is compact in the weak* topology [11, p. 19].

Now, since $\mathcal{F}_{\mathbf{c}_{0:n}}$ is a compact convex set in a locally convex topological linear space, it is the closure of the convex hull of its extreme points [11, p. 28-29]. *Extreme points* have the property that they are not a (nontrivial) convex combination of elements in the set. The set of extreme points, i.e. the *essential boundary*, will be denoted by $\mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$ and its characterization will be given shortly.

Finally, if δ is a weak*-continuous and jointly convex function, then the diameter is attained as the precise distance between two elements in $\mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$. The set of extreme points admits a finite dimensional characterization and hence, computation of the diameter becomes a tractable problem.

Proposition 4: Let $\mathbf{c}_{0:n}$ be a nonnegative sequence. Then, $d\mu \in \mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$ if and only if $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$ and the support of $d\mu$ consists of at most $2n+1$ points.

Proof: See appendix. ■

The proposition states that any $d\mu \in \mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$ is a singular measure with at most $(2n+1)$ points of increase. Hence, to specify $\mathbf{c}$ we only need to specify $(c_{n+1}, c_{n+2}, \ldots, c_{2n+1})$ so that $T_{2n+1} \geq 0$ and singular. Thus, $\mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$ admits a finite dimensional characterization and $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ reduces to a finite dimensional problem.

V. Distance functions

In the previous sections we argued that weak* continuity and joint convexity are desirable properties of any distance function between nonnegative measures. Three possible choices are given below.

(i) If we consider measures as functionals on $C(\mathbb{T})$, it is natural to quantify distance between such based on their action on particular subsets of $C(\mathbb{T})$, e.g.,

$$\delta(d\mu_1, d\mu_2) = \sup_{z \in K} \left| \frac{1}{2\pi} \int_{\mathbb{T}} g_z d\mu_1 - \frac{1}{2\pi} \int_{\mathbb{T}} g_z d\mu_2 \right|$$

with $\{g_z\}_{z \in K} \subset C(\mathbb{T})$ and K being an indexing set. For this to be a metric, the set $\{g_z\}_{z \in K}$ needs to be sufficiently rich to separate measures. An interesting special case is when K is taken to be a compact subset of $\mathbb{D}$ and g_z the Poisson kernel. In this case the distance reduces to the sup-norm over K between harmonic functions on the disc

$$\delta_K(d\mu_1, d\mu_2) = \max_{z \in K} |P[d\mu_1 - d\mu_2](z)|.$$

If $K \subset \mathbb{D}$ has a nonempty interior δ_K defines a metric which is treated in the next section.

(ii) Alternative possibilities include L_1-distances between respective functionals over a set of functions

$$\delta(d\mu_1, d\mu_2) = \int_{z \in K} \left| \frac{1}{2\pi} \int_{\mathbb{T}} g_z d\mu_1 - \frac{1}{2\pi} \int_{\mathbb{T}} g_z d\mu_2 \right| dz,$$

and in particular, the $L^1(\mathbb{D})$-distance between the respective harmonic functions on the disc

$$\delta_1(d\mu_1, d\mu_2) = \int_{\mathbb{D}} |P[d\mu_1 - d\mu_2](z)| dx dy, \; z = x + iy.$$

(iii) A range of possibilities opens up if we forgo the requirement of δ being a metric, and are prepared to consider pseudodistances, such as a relative entropy functional (e.g., see [9]). For the purposes of this paper, the usual expression for the relative entropy is not applicable because it is not weak* continuous. However, alternative choices are possible as for instance

$$\delta(d\mu_1, d\mu_2) = \int_{\mathbb{D}} P[d\mu_1] \log\left(\frac{P[d\mu_1]}{P[d\mu_2]}\right) dx dy, \; z = x + iy.$$

This notion of distance is again weak* continuous and jointly convex, and will be discussed in [14].

VI. The case ρ_{δ_K}

The size of the uncertainty set with respect to the distance δ_K turns out to be especially easy to compute. Indeed, the diameter is attained on a special subset of the essential boundary which corresponds to measures with only $n+1$ points of increase (i.e., support). This is the content of the following proposition.

Proposition 5: Let $\mathbf{c}_{0:n}$ be a positive covariance sequence and let $K \subset \mathbb{D}$ be closed. Then

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}}) = \max_{z\in K}\left\{2\left(\left|\frac{\frac{2}{1-z\bar z}+(b_z,d_z)}{(b_z,b_z)}\right|^2-\frac{(d_z,d_z)}{(b_z,b_z)}\right)^{\frac{1}{2}}\right\},$$

where

$$b_z=\begin{pmatrix} z^{-1}\\ z^{-2}\\ \vdots\\ z^{-n-1}\end{pmatrix},\ d_z=\begin{pmatrix} z^{-1}(c_0)\\ z^{-2}(c_0+2c_1z)\\ \vdots\\ z^{-n-1}(c_0+2c_1z+\cdots+2c_nz^n)\end{pmatrix},$$

and (x,y) denote the inner product $y^*T_n^{-1}x$. Furthermore, $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ is attained as the distance between two elements of $\mathcal{F}_{\mathbf{c}_{0:n}}$ which are both singular with support containing at most $n+1$ points.

Proof: See appendix. ■

Both claims in Proposition 5 can be used separately for computing $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$. The first one suggests finding the maximum of a real valued function over K. The second claim suggests search for a maximum for $\delta_K(d\mu_1,d\mu_2)$ over a rather small subset of $\mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$, namely nonnegative sequences $\mathbf{c}_{0:(n+1)}$ parametrized by c_{n+1} being a solution of the quadratic equation

$$\det(T_{n+1})=0.$$

The corresponding values for c_{n+1} lie on a circle in the complex plane, and hence, computation of $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ will require search on a torus (each of the two extremal $d\mu_1$, $d\mu_2$ where the diameter is attained can be thought of as points on the circle).

As an example, Figure 1 shows $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ for

$$\mathbf{c}_{0:2}=(1,\ c_1,\ c_2)$$

as a function of the corresponding Schur parameters [10]

$$\begin{array}{rcl} -1< & \gamma_1:=c_1 & <1,\\ -1\le & \gamma_2:=\dfrac{\det\begin{pmatrix} c_1 & c_2\\ 1 & c_1\end{pmatrix}}{\det\begin{pmatrix} 1 & c_1\\ \bar c_1 & 1\end{pmatrix}} & \le 1,\end{array}$$

and K is taken as $\{z\ :\ |z|\le 0.5\}\subset\mathbb{D}$.

The plot confirms that the diameter decreases to zero as the parameters or, alternatively, the covariances c_1 and c_2, tend to the boundary of the "positive" region (which in the Schur coordinates corresponds to the unit square). However, it is interesting to note that the diameter of $\mathcal{F}_{\mathbf{c}_{0:n}}$ as a function of $\mathbf{c}_{0:n}$ has several local maxima.

Conclusions

The purpose of this paper has been to bring attention to the issue of quantifying uncertainty in the context of moment problems. Such problems arise in a variety of engineering applications (signal processing, feedback control, circuit theory, theory of measurements, statistical mechanics, etc., see [2], [3], [6], [7], [8]). To this end we seek a distance function with

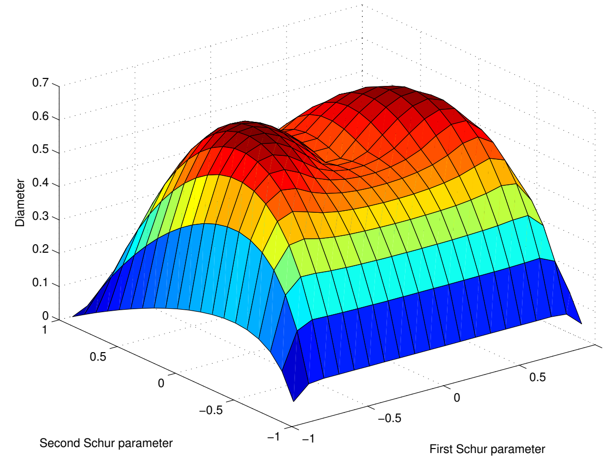

Fig. 1. ρ_{δ_K} *as a function of* γ_1,γ_2 *when* $c_0=1$. $K=\{z:|z|\le 0.5\}$

certain desirable properties: first that the induced diameter of the uncertainty set reduces to zero as the set reduces to a singleton, and second that the diameter can be computed with reasonable efficiency. Certain alternatives are proposed which are further explored and discussed in [14].

VII. Acknowledgements

The authors would like to thank Professor A. Lindquist for his support and many inspiring conversations.

VIII. Appendix: background & technical arguments

Orthogonal polynomials and Schur coefficients

Let $\mathbf{c}$ be a nonnegative covariance sequence with corresponding measure $d\mu$ and consider the inner product

$$\langle a(z),b(z)\rangle=\frac{1}{2\pi}\int_{\mathbb{T}}a(e^{i\theta})\overline{b(e^{i\theta})}d\mu(\theta).$$

The so-called *orthogonal polynomials (of the first kind)* $\phi_k(z)$ [10] are (uniquely defined) monic polynomials with $\deg\phi_k(z)=k$, $k=0,1\ldots$, which are orthogonal with respect to $\langle\cdot,\cdot\rangle$. They are shown [10] to satisfy the recursion

$$\begin{array}{rcl}\phi_{k+1}(z) &=& z\phi_k(z)-\bar\gamma_k\phi_k(z)^*,\\ \phi_{k+1}(z)^* &=& \phi_k(z)^*-z\gamma_k\phi_k(z),\end{array}\qquad(2)$$

where $\phi_k(z)^*=z^k\overline{\phi_k(\bar z^{-1})}$ and $\{\gamma_k\}_{k=1}^\infty$ are the so-called *Schur parameters.*

The *orthogonal polynomials of the second kind* are defined by

$$\psi_k(z)=\frac{1}{c_0}[(\overline{f(\bar z^{-1})})\phi_k(z)]_+,$$

where $[\cdot]_+$ denote "the polynomial part of". They are also "orthogonal polynomials" but with respect to a certain "inverted" covariance (corresponding to the negative of the original Schur parameters, cf. [10]) and satisfy the recursion

$$\begin{array}{rcl}\psi_{k+1}(z) &=& z\psi_k(z)+\bar\gamma_k\psi_k(z)^*,\\ \psi_{k+1}(z)^* &=& \psi_k(z)^*+z\gamma_k\psi_k(z).\end{array}\qquad(3)$$

The positive-real function $f(z) = H[d\mu](z)$ may be expressed using the orthogonal polynomials as

$$f(z) = c_0 \frac{\psi_k(z)^* + zs_{k+1}(z)\psi_k(z)}{\phi_k(z)^* - zs_{k+1}(z)\phi_k(z)},$$

where s_{k+1} belong to the Schur class $\mathcal{S}$, i.e. the class of analytic functions on $\mathbb{D}$ uniformly bounded by 1. Equations (2-3) lead to

$$f(z) = c_0 \frac{1 + zs_1(z)}{1 - zs_1(z)}, \quad s_k(z) = \frac{\gamma_k + zs_{k+1}(z)}{1 + z\bar{\gamma}_k s_{k+1}(z)}, \qquad (4)$$

for $k = 1, 2, \ldots$.

Lemma 1: Let $f_j(z) = H[d\mu_j](z)$ where $d\mu_j$ are non-negative measures, $j = 1, 2$. Let $\{\gamma_k^j\}$, $\{s_k^j(z)\}$ be the corresponding sequences of Schur parameters and Schur functions. Then the following inequalities hold pointwise for $z \in \mathbb{D}$:

$$|f_1 - f_2| \leq \frac{2|c_0^1 - c_0^2| + |c_0^1 + c_0^2||s_1^1 - s_1^2|}{(1 - |z|)^2}, \text{ and}$$

$$|s_k^1 - s_k^2| \leq \frac{6|\gamma_k^1 - \gamma_k^2| + |s_{k+1}^1 - s_{k+1}^2|\sqrt{(1-|\gamma_k^1|)(1-|\gamma_k^2|)}}{(1 - |z|)^2}.$$

Proof: They follow easily from (4). ■

Proof: [Proposition 3] Let $d\mu$ being the unique measure specified by $\hat{\mathbf{c}}_{0:n}$, and let $d\mu_\ell \in \mathcal{F}_{\mathbf{c}_{0:n}(\ell)}$ denote a sequence of measures, for $\ell = 1, 2, \ldots$. The Schur parameters γ_k corresponding to $d\mu$ satisfies $|\gamma_k| < 1$ for $k = 1, \ldots, m-1$, $|\gamma_m| = 1$, for some $m \leq n$. Since $c_k(\ell) \to \hat{c}_k$, the k:th Schur parameter corresponding to $d\mu_\ell$ converges to γ_k, for $k = 1, \ldots, m$ [10, §8.2, §8.5]. By Lemma 1 the harmonic functions representing $P[d\mu_\ell](z)$ converge to $P[d\mu](z)$ uniformly on compact subsets of $\mathbb{D}$, hence (Proposition 1) $d\mu_\ell \to d\mu$ weak*. Weak* continuity of δ now ensures that the supremum of $\delta(d\mu_\ell, d\mu)$ over $d\mu_\ell \in \mathcal{F}_{\mathbf{c}_{0:n}(\ell)}$ goes to zero with n, and so does $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}(\ell)})$. ■

Proof: [Proposition 4] Assume that $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$ has more than $2n+1$ points of increase. Then let $[E_1, E_2, ..., E_{2n+1}, E_{2n+2}]$ be a partition of $\mathbb{T}$ such that $\mu(E_j) > 0$. Denote $\mu_j = \mu|_{E_j}$ and denote by $\mathbf{c}_{0:n}(j)$ the covariances corresponding to $d\mu_j$. By assumption $c_k = \frac{1}{2\pi}\int e^{-i\theta k} d\mu$. It follows that the set of linear equations

$$\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} c_0(1) & \cdots & c_0(2n+2) \\ c_1(1) & \cdots & c_1(2n+2) \\ \vdots & & \vdots \\ c_n(1) & \cdots & c_n(2n+2) \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{2n+2} \end{bmatrix}$$

has real positive solution $\tau_i \geq 0$ $(i = 1, 2, \ldots, 2n+2)$. Dimensionality considerations show that the solution is not unique. Indeed the matrix $[c_k(j)]$ has range with real dimension $2n+1$ whereas the domain has dimension $2n+2$. Hence there is at least a one-dimensional family of solutions $(\tau_1, \tau_2, \ldots, \tau_{2n+2})$. Since $d\mu$ corresponds to a positive solution, $d\mu$ belongs to the relative interior of this family, hence it can be written as a proper convex combination of two other measures in $\mathcal{F}_{\mathbf{c}_{0:n}}$, i.e. $d\mu \notin \mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$.

To show the converse, let $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$ be a measure with at most $2n+1$ points of increase. Then it is of the form $d\mu_\tau = \sum_{j=1}^{2n+1} \tau_j d\mu_j$, where $d\mu_j$ is the measure with unit mass at t_j and all t_j are distinct. Denote by $c_k(j)$ the covariances corresponding to $d\mu_j$. Then τ_j satisfies

$$\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} c_0(1) & \cdots & c_0(2n+1) \\ c_1(1) & \cdots & c_1(2n+1) \\ \vdots & & \vdots \\ c_n(1) & \cdots & c_n(2n+1) \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{2n+1} \end{bmatrix}. \quad (5)$$

The linear mapping $[c_k(j)]$ is nonsingular and have both range and domain of real dimension 2n+1. Therefore τ is the unique real solution of Equation 5. If $d\mu$ is a convex combination of $d\nu_\ell \in \mathcal{F}_{\mathbf{c}_{0:n}}$, $\ell = 1, 2$, then $\mathrm{supp}(d\nu_1) \cup \mathrm{supp}(d\nu_2) \subset \mathrm{supp}(d\mu)$, hence they are of the form $d\nu_\ell = \sum_{j=1}^{2n+1} \tau_j(\ell) d\mu_j$, $\tau_j(\ell)$ real. But since $d\mu$ is the only measure of this form in $\mathcal{F}_{\mathbf{c}_{0:n}}$ we have $d\mu = d\nu_1 = d\nu_2$, hence $d\mu \in \mathrm{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$. ■

Proof: [Proposition 5] There exists an analytic function $f(z) = H[d\mu](z), d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$, such that $f(z) = w_z$ if and only if its associated Pick matrix is nonnegative [15], i.e.

$$\begin{pmatrix} 2T_n & b_z w_z - d_z \\ \bar{w}_z b_z^* - d_z^* & \frac{w_z + \bar{w}_z}{1 - z\bar{z}} \end{pmatrix} \geq 0. \qquad (6)$$

By using Schur's lemma and completing the squares we arrive at

$$\left| w_z - \frac{\frac{2}{1-z\bar{z}} + (d_z, b_z)}{(b_z, b_z)} \right|^2 \leq \left| \frac{\frac{2}{1-z\bar{z}} + (b_z, d_z)}{(b_z, b_z)} \right|^2 - \frac{(d_z, d_z)}{(b_z, b_z)}, \quad (7)$$

where equality holds if and only if the Pick matrix (6) is singular. From this, the first part of Proposition 5 follows. Since the maximum is obtained when equality holds in Equation 7, the associated Pick matrices are singular. Hence the solutions are unique and correspond to measures with support on $n+1$ points [5, prop. 2]. ■

References

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

[2] 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.

[3] 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.

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

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

[6] 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.

[7] T.T. Georgiou, "Solution of the general moment problem via a one-parameter imbedding," *IEEE Trans. on Automatic Control*, **50(6)**: 811-826, June 2005.

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

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

[10] Ya. L. Geronimus, **Orthogonal Polynomials**, Consultants Bureau, 210 pages, 1961.

[11] I.V. Girsanov, **Lectures on Mathematical Theory of Extremum Problems**, Springer, 1972.

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

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

[14] J. Karlsson and T.T. Georgiou, "Quantifying uncertainty in the moment problem," report in preparation, KTH.

[15] I.V. Kovalishina and V.P. Potapov, **Integral Representation of Hermitian Positive Functions,**, Khark'hov Railway Engineering Inst., Khar'kov 2001. English translation by T. Ando, Sapporo, Japan, 1981.

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

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

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