The Uncertain Generalized Moment Problem with Complexity Constraint

Christopher I. Byrnes[1] and Anders Lindquist[2]

[1] Washington University, St. Louis, MO, USA, chrisbyrnes@seas.wustl.edu
[2] Royal institute of Technology, Stockholm, Sweden, alq@math.kth.se

This paper[3] is dedicated to Arthur Krener – a great researcher, a great teacher and a great friend – on the occasion of his 60th birthday. In this work we study the generalized moment problem with complexity constraints in the case where the actual values of the moments are uncertain. For example, in spectral estimation the moments correspond to estimates of covariance lags computed from a finite observation record, which inevitably leads to statistical errors, a problem studied earlier by Shankwitz and Georgiou. Our approach is a combination of methods drawn from optimization and the differentiable approach to geometry and topology. In particular, we give an intrinsic geometric derivation of the Legendre transform and use it to describe convexity properties of the solution to the generalized moment problems as the moments vary over an arbitrary compact convex set of possible values. This is also interpreted in terms of minimizing the Kullback-Leibler divergence for the generalized moment problem.

1 Introduction

Let $\alpha_0, \alpha_1, \cdots, \alpha_n$ be a sequence of C^2 functions defined on some interval $\mathcal{I}$ of the real line. Given a suitable sequence of complex numbers, $c_0, c_1, \cdots, c_n$, we are interested in moment problems of the form

$$\int_{\mathcal{I}} \alpha_k(t)\Phi(t)dt = c_k, \quad k = 0, 1, \cdots, n, \tag{1}$$

for functions $\Phi \in L^1_+(\mathcal{I})$, where $L^1_+(\mathcal{I})$ is the space of positive functions in $L^1(\mathcal{I})$.

In fact, suppose that $\alpha_0, \ldots, \alpha_{r-1}$ are real functions and $\alpha_r, \ldots, \alpha_n$ are complex-valued functions whose real and imaginary parts, taken together with

[3] This research was supported in part by grants from AFOSR, VR, Institut Mittag-Leffler, and SBC.

$\alpha_0, \dots, \alpha_{r-1}$ are linearly independent over $\mathbb{R}$. This is no restriction since a purely imaginary moment condition can always be reduced to a real one. For simplicity of exposition, we also assume that $\alpha_0 = 1$. Let $\mathfrak{P}$ be the real vector space that is the sum of the real span of $\alpha_0, \dots, \alpha_{r-1}$ and the complex span of $\alpha_r, \dots, \alpha_n$. Hence, the real dimension of $\mathfrak{P}$ is $2n - r + 2$. If $\mathfrak{P}_+$ denotes the subset of all functions in $\mathfrak{P}$ that have a positive real part on $\mathcal{I}$, then $\mathfrak{P}_+$ is a nonempty, open, convex subset of dimension $2n - r + 2$.

For this moment problem to have a solution it is clearly necessary that the sequence $c_0, c_1, \cdots, c_n$ is *positive* in the sense that

$$\langle c, q \rangle := \mathrm{Re} \sum_{k=0}^{n} q_k c_k > 0 \tag{2}$$

for all $(q_0, q_1, \cdots, q_n) \in \mathbb{R}^r \times \mathbb{C}^{n-r+1}$ such that

$$q := \sum_{k=0}^{n} q_k \alpha_k \in \overline{\mathfrak{P}}_+ \smallsetminus \{0\}, \tag{3}$$

where $\overline{\mathfrak{P}}_+$ is the closure of $\mathfrak{P}_+$. Indeed,

$$\langle c, q \rangle = \int_{\mathcal{I}} \left[\mathrm{Re} \sum_{k=0}^{n} q_k \alpha_k \right] \Phi dt > 0,$$

whenever (3) holds. If $\mathfrak{C}_+$ denotes the space of positive sequences, then $\mathfrak{C}_+$ is a nonempty, open, convex subset of dimension $2n - r + 2$.

In [3] we considered the problem to find, for each Ψ in some class $\mathcal{G}_+$, the particular solution Φ to the moment problem (1) minimizing the Kullback-Leibler divergence

$$\mathbb{I}_{\Psi}(\Phi) = \int_{\mathcal{I}} \Psi(t) \log \frac{\Psi(t)}{\Phi(t)} dt. \tag{4}$$

Here $\mathcal{G}_+$ is the class of functions in $L^1_+(\mathcal{I})$ satisfying the normalization condition

$$\int_{\mathcal{I}} \Psi(t) dt = 1 \tag{5}$$

and the integrability conditions

$$\left| \int_{\mathcal{I}} \alpha_k \frac{\Psi}{\mathrm{Re}\{q\}} \, dt \right| < \infty, \quad k = 0, 1, \dots, n, \tag{6}$$

for all $q \in \mathfrak{P}_+$. If $\mathcal{I}$ is a finite interval, (6) of course holds for all $\Psi \in L^1_+(\mathcal{I})$.

In fact, Ψ could be regarded as some *a priori* estimate, and, as was done in [10] for spectral densities, we want to find the function Φ that is "closest" to Ψ in the Kullback-Leibler distance and also satisfies the moment conditions (1). This notion of distance arises in many applications, e.g., in coding theory [8] and probability and statistics [13, 11, 9]. Note, however, that Kullback-Leibler divergence is not really a metric, but, if we normalize by taking $c_0, c_1, \cdots, c_n$ in

$$\mathcal{C}_+ := \{c \in \mathfrak{C}_+ \mid c_0 = 1\} \tag{7}$$

so that Φ satisfies (5), the Kullback-Leibler divergence (4) is nonnegative, and it is zero if and only if $\Phi = \Psi$.

In [3] we proved that the problem to minimize (4), subject to the moment conditions (1), has a unique solution for each $\Psi \in \mathcal{G}_+$ and $c \in \mathfrak{C}_+$ and that this solution has the form

$$\Phi(t) = \frac{\Psi(t)}{\mathrm{Re}\{q(t)\}} \tag{8}$$

for some $q \in \mathfrak{P}_+$, which can be determined as the unique minimum in $\mathfrak{P}_+$ of the strictly convex functional

$$\mathbb{J}_\Psi(q) = \langle c, q \rangle - \int_{\mathcal{I}} \Psi \log\left(\mathrm{Re}\{q(t)\}\right) dt. \tag{9}$$

This ties up to a large body of literature [4, 7, 5, 6, 2, 3, 10] dealing with interpolation problems with complexity constraints.

In this paper we consider a modified optimization problem in which c is allowed to vary in some compact, convex subset $\mathcal{C}_0$ of $\mathcal{C}_+$, where $\mathcal{C}_+ \subset \mathfrak{C}_+$ is given by (7). In fact, the moments $c_1, c_2, \cdots, c_n$ may not be precisely determined, but only known up to membership in $\mathcal{C}_0$. The problem at hand is then

Problem 1. Find a pair $(\Phi, c) \in L^1_+(\mathcal{I}) \times \mathcal{C}_0$ that minimizes the Kullback-Leibler divergence (4) subject to the moment conditions (1).

We will show that this problem has a unique minimum and that the corresponding c lies in the interior of $\mathcal{C}_0$ only if Ψ satisfies the moment conditions, in which case the optimal Φ equals Ψ.

An important special case of Problem 1 was solved in [15]. In [15] the uncertain covariance extension problem, as a tool for spectral estimation, is noted to have two fundamentally different kinds of uncertainty. It is now known [3, 10] that the rational covariance extension problem can be solved by minimizing the Kullback-Leibler divergence (4), where Ψ is an arbitrary positive trigonometric polynomial of degree at most n, and where the functions $\alpha_0, \alpha_1, \dots, \alpha_n$ are the trigonometric monomials, i.e, $\alpha_k(t) = e^{ikt}$, $k = 0, 1, \dots, n$. The corresponding moments are then the covariance lags of an underlying process.

The uncertainty involving the choice of Ψ is resolved in [15] by choosing $\Psi = 1$. Then minimizing the Kullback-Leibler divergence is equivalent to finding the maximum-entropy solution, corresponding to having no *a priori* information about the estimated process, namely the solution to the trigonometric moment problem maximizing the entropy gain

$$\int_{\mathcal{I}} \log \Phi(t) dt. \tag{10}$$

The other fundamental uncertainty in this problem arises from the statistical errors introduced in estimating the covariance lags from a given finite observation record. This was modeled in [15] by assuming that the true covariance lags are constrained to lie in the polyhedral set

$$c_k \in [c_k^-, c_k^+], \quad k = 0, 1, \dots, n. \tag{11}$$

In this setting, it is shown that the maximal value of (10) subject to the moment conditions is a strictly convex function on the polytope $\mathcal{C}_0$ defined by (11) and hence that there is a unique choice of $c \in \mathcal{C}_0$ maximizing the entropy gain. As will be shown in this paper, this is a special case of our general solution to Problem 1.

2 Background

The problem described above is related to a moment problem with a certain complexity constraint: In [2, 3] we proved that the moment problem (1) with the complexity constraint (8) has a unique solution. More precisely, we proved

Theorem 1. *For any $\Psi \in \mathcal{G}_+$ and $c \in \mathfrak{C}_+$, the function $F : \mathfrak{P}_+ \to \mathfrak{C}_+$, defined componentwise by*

$$F_k(q) = \int_{\mathcal{I}} \alpha_k(t) \frac{\Psi(t)}{\text{Re}\{q(t)\}} dt, \quad k = 0, 1, \dots, n, \tag{12}$$

is a diffeomorphism. In fact, the moment problem (1) *with the complexity constraint* (8) *has a unique solution $\hat{q} \in \mathfrak{P}_+$, which is determined by c and Ψ as the unique minimum in $\mathfrak{P}_+$ of the strictly convex functional* (9)*.*

Note that $\mathbb{J}_\Psi(q)$ is finite for all $q \in \mathfrak{P}_+$. In fact, by Jensen's inequality,

$$-\log \int_{\mathcal{I}} \frac{\Psi}{\text{Re}\{q(t)\}} dt \leq \int_{\mathcal{I}} \Psi \log\left(\text{Re}\{q(t)\}\right) dt \leq \log \int_{\mathcal{I}} \text{Re}\{q(t)\} \Psi dt,$$

where both bounds are finite by (6). (To see this, for the lower bound take $k = 0$; for the upper bound first take $q = 1$ in (6), and then form the appropriate linear combination.) In this paper we shall give a new proof of Theorem 1 by using methods from convex analysis.

As proved in [3], following the same pattern as in [4, 5, 6], the optimization problem of Theorem 1 is the dual problem in the sense of mathematical programming of the constrained optimization problem in the following theorem.

Theorem 2. *For any choice of $\Psi \in \mathcal{G}_+$, the constrained optimization problem to minimize the Kullback-Leibler divergence* (4) *over all $\Phi \in L^1_+(\mathcal{I})$ subject to the constraints* (1) *has unique solution $\hat{\Phi}$, and it has the form*

$$\hat{\Phi} = \frac{\Psi}{\mathrm{Re}\{\hat{q}\}},$$

where $\hat{q} \in \mathfrak{P}_+$ is the unique minimizer of (9)*. Moreover, for all $\Phi \in L^1_+(\mathcal{I})$ and $q \in \mathfrak{P}_+$,*

$$-\mathbb{I}_\Psi(\Phi) \leq \mathbb{J}_\Psi(q) - 1 \tag{13}$$

with equality if and only if $q = \hat{q}$ and $\Phi = \hat{\Phi}$.

3 The uncertain moment problem

We are now a position to solve Problem 1. We shall need the following definition [14, p. 251].

Definition 1. A function f is *essentially smooth* if

1. $\mathrm{int}(\mathrm{dom} f)$ is nonempty;
2. f is differentiable throughout $\mathrm{int}(\mathrm{dom} f)$;
3. $\lim_{k\to\infty} |\nabla f(x^{(k)})| = +\infty$ whenever $\{x^{(k)}\}$ is a sequence in $\mathrm{int}(\mathrm{dom} f)$ converging to the boundary of $\mathrm{int}(\mathrm{dom} f)$.

An essentially smooth function such that $\mathrm{int}(\mathrm{dom} f)$ is convex and f is a strictly convex function on $\mathrm{int}(\mathrm{dom} f)$ is is called a *convex function of Legendre type.*

The optimal point $\mathbb{I}_\Psi(\hat{\Phi})$ of Theorem 2 clearly depends on c, and hence we may define a function

$$\varphi : \mathfrak{C}_+ \to \mathbb{R}$$

which sends c to $\mathbb{I}_\Psi(\hat{\Phi})$, i.e.,

$$c \mapsto \int_{\mathcal{I}} \Psi(t) \log \frac{\Psi(t)}{\hat{\Phi}(t)} dt. \tag{14}$$

We also write $\hat{q}(c)$ to emphasize that the unique minimizer $\hat{q}$ in $\mathfrak{P}_+$ of the functional (9) depends on c. Similarly, we write $\hat{\Phi}(c)$ for the unique minimizer of Theorem 2. Then, by Theorem 2,

$$\hat{\Phi}(c) = \frac{\Psi}{\mathrm{Re}\{\hat{q}\}(c)}, \quad \text{for all } c \in \mathfrak{C}_+.$$

Theorem 3. *The function φ is a convex function of Legendre type. In particular, φ is strictly convex, and the problem to minimize φ over the compact, convex subset $\mathcal{C}_0$ of $\mathcal{C}_+$ has a unique solution. The minimizing point $\hat{c}$ belongs to the interior of $\mathcal{C}_0$ only if Ψ satisfies the moment conditions* (1)*, in which case*

$$\hat{q}(\hat{c}) = 1.$$

The gradient of φ is given by

$$\nabla\varphi(c) = -\hat{q}(c), \tag{15}$$

and the Hessian is the inverse of the matrix

$$H(c) := \left[\int_{\mathcal{I}} \alpha_j(t) \frac{\Psi(t)}{\left(\mathrm{Re}\{\hat{q}(c)(t)\}\right)^2} \alpha_k(t) dt\right]_{j,k=0}^{n}. \tag{16}$$

The proof of this theorem will be given in Section 5.

As an illustration, we can use Newton's method to solve Problem 1. In fact, suppose that $\mathcal{C}_0$ has an nonempty interior. Then, for any $c^{(0)}$ in the interior of $\mathcal{C}_0$, the recursion

$$c^{(\nu+1)} = c^{(\nu)} + \lambda_\nu \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} H(c^{(\nu)}) \hat{q}(c^{(\nu)}), \quad \nu = 0, 1, 2, \dots \tag{17}$$

will converge to $\hat{c}$ for a suitable choice of $\{\lambda_\nu\}$ keeping the sequence inside $\mathcal{C}_0$. This algorithm can be implemented in the following way. For $\nu = 0, 1, 2, \dots,$ the gradient $\hat{q}(c^{(\nu)})$ is determined as the unique minimum in $\mathfrak{P}_+$ of the strictly convex functional

$$\mathbb{J}_\Psi^{(\nu)}(q) = \langle c^{(\nu)}, q\rangle - \int_{\mathcal{I}} \Psi \log\left(\mathrm{Re}\{q(t)\}\right) dt, \tag{18}$$

and then $c^{(\nu+1)}$ is obtained from (17).

As an example, consider the special, but important, case that $\mathcal{C}_0$ is defined as the polyhedral set of all $c = (c_0, c_1, \cdots, c_n) \in \mathcal{C}_+$ satisfying (11). The Lagrange relaxed problem is then to minimize

$$L(c, \lambda^-, \lambda^+) = \varphi(c) + \sum_{k=0}^{n} \lambda^-(c_k - c_k^-) + \sum_{k=0}^{n} \lambda^+(c_k^+ - c_k), \tag{19}$$

where $\lambda_k^- \geq 0$ and $\lambda_k^+ \geq 0$, $k = 0, 1, \dots, n$, are Lagrange multipliers. By Theorem 3, the Lagrangian has a unique stationary point that satisfies

$$\hat{q}(c) = \lambda^+ - \lambda^-. \tag{20}$$

By the principle of complementary slackness, a Lagrange multiplier can be positive only when the corresponding constraint is satisfied with equality at the optimum. In particular, if all components of $\hat{q}(\hat{c})$ are nonzero, $\hat{c}$ must be a corner point of the polyhedral set $\mathcal{C}_0$.

4 A derivation of the Legendre transform from a differentiable viewpoint

Suppose U is an open subset of $\mathbb{R}^N$, which is diffeomorphic to $\mathbb{R}^N$, and that F is a C^1 map

$$F : U \to \mathbb{R}^N$$

with a Jacobian, $\mathrm{Jac}_q(F)$, which is invertible for each $q \in U$. A useful formulation of the Poincaré Lemma is that $\mathrm{Jac}_q(F)$ is symmetric for each $q \in U$ if and only if F is the gradient vector, ∇f, for some C^2 function

$$f : U \to \mathbb{R},$$

which is unique up to a constant of integration.

Remark 1. Here, we mean symmetric when represented as a matrix in the standard basis of $\mathbb{R}^N$, i.e., symmetric as an operator with respect to the standard inner product. We interpret the gradient as a column vector using this inner product as well.

Alternatively, consider the 1-form

$$\omega = \sum_{k=1}^{N} F_k dq_k,$$

where F_k and q_k denote the kth component of F and q, respectively. To say that $\mathrm{Jac}_q(F)$ is symmetric for all $q \in U$ is to say that $d\omega = 0$ on U, and therefore $\omega = df$ for an f as above.

More generally,

$$\sum_{k=1}^{N} \left(F_k dq_k - q_k^* dq_k\right) = df(q) - \sum_{k=1}^{N} q_k^* dq_k$$

so that

$$df(q) = \sum_{k=1}^{N} q_k^* dq_k \quad \Leftrightarrow \quad F(q) = q^* \quad \Leftrightarrow \quad \nabla f(q) = q^*. \tag{21}$$

We now specialize to the strictly convex case, i.e., we suppose that U is convex and that $\mathrm{Jac}_q(F)$ is positive definite for all $q \in U$. Alternatively, we could begin our construction with a strictly convex C^2 function f. In this case, we note that (21) is equivalent to

$$\inf_p \{f(p) - \langle p, q^* \rangle\} = f(q) - \langle q, q^* \rangle \quad \Leftrightarrow \quad \nabla f(q) = q^*. \tag{22}$$

The left hand side of the equivalence (22) defines a function of q^*, which we denote by $g(q^*)$, and which we will soon construct in an intrinsic, geometric fashion. For now, it suffices to note that, in light of (21), we obtain the following expression for g:

$$g(q^*) = f((\nabla f)^{-1}(q^*)) - \langle q^*, (\nabla f)^{-1}(q^*)\rangle. \tag{23}$$

In fact, since f is strictly convex, the map F is injective. Since F has an everywhere nonvanishing Jacobian, by the inverse function theorem, F is a diffeomorpism between U and $V := F(U)$, where V is an open subset of $\mathbb{R}^N$. Since the inverse of a positive definite matrix is positive definite, F^{-1} has an everywhere nonsingular symmetric Jacobian, $\mathrm{Jac}_{q^*}(F^{-1})$. Therefore, we may apply our general construction to find, up to a constant of integration, a unique C^2 function

$$f^* : V \to \mathbb{R}$$

satisfying

$$\sum_{k=1}^{N} [F^{-1}]_k dq_k^* = df^*(q_k^*)$$

and, more generally,

$$\sum_{k=1}^{N} \left([F^{-1}]_k dq_k^* - q_k dq_k^*\right) = df^*(q_k^*) - \sum_{k=1}^{N} q_k dq_k^*$$

and consequently

$$df^*(q^*) = \sum_{k=1}^{N} q_k dq_k^* \quad \Leftrightarrow \quad F^{-1}(q^*) = q \quad \Leftrightarrow \quad \nabla f^*(q^*) = q. \tag{24}$$

Of course, this geometric duality has several corollaries. Fix $q_0 \in U$ and $q_0^* := F(q_0) \in V$. Let $\hat{q}$ be an arbitrary point in U and denote its image, $F(\hat{q})$, by $\hat{q}^*$. Let γ be any smooth oriented curve starting at q_0 and ending at $\hat{q}$, and consider $\gamma^* := F(\gamma)$. We may then compute the following path integral as a function of the upper limit,

$$f^*(\hat{q}^*) - f^*(q_0^*) = \int_{\gamma^*} df^*(q^*) = \int_{\gamma^*} \sum_{k=1}^{N} [F^{-1}]_k(q^*) dq_k^* = \int_{\gamma} \sum_{k=1}^{N} q_k dF_k. \tag{25}$$

Then, integrating by parts, we obtain

$$\begin{aligned} f^*(\hat{q}^*) &= f^*(q_0^*) + \sum_{k=1}^{N} q_k F_k \Big|_{q_0}^{\hat{q}} - \sum_{k=1}^{N} \int_{\gamma} F_k dq_k \\ &= f^*(q_0^*) + \langle q, \nabla f\rangle \Big|_{q_0}^{\hat{q}} - \int_{\gamma} df \\ &= \langle \hat{q}, \nabla f(\hat{q})\rangle - f(\hat{q}) + \kappa, \end{aligned}$$

where

$$\kappa := f^*(q_0) - \langle q_0, \nabla f(q_0)\rangle + f(q_0)$$

is a constant of integration for f^*, which we may set equal to zero. Therefore, since $\hat{q} = (\nabla f)^{-1}(\hat{q}^*)$ and $\hat{q}^* = \nabla f(\hat{q})$,

$$f^*(\hat{q}^*) = \langle (\nabla f)^{-1}(\hat{q}^*), \hat{q}^* \rangle - f\big((\nabla f)^{-1}(\hat{q}^*)\big),$$

or, recalling that $\hat{q}$ is arbitrary,

$$f^*(q^*) = \langle (\nabla f)^{-1}(q^*), q^* \rangle - f\big((\nabla f)^{-1}(q^*)\big) = -g(q^*).$$

Remark 2. Since our fundamental starting point assumes that F has a symmetric everywhere nonsingular Jacobian, the above analysis extends to strictly concave functions, the only change being that the infima be replaced by suprema. Furthermore, since the Hessian of f^* is the inverse of the Hessian of f, it follows that, on any open convex subset of V, f^* will be strictly convex (strictly concave) whenever f is strictly convex (strictly concave).

Remark 3. These expressions are well-known in convex optimization theory. (See, e.g., [12, 14].) Indeed, since $f^* = -g$, (22) yields

$$f^*(q^*) = \sup_{q \in U} \left\{ \langle q^*, q \rangle - f(q) \right\}, \tag{26}$$

which is referred to as the *conjugate function* of f. Then, (23) yields

$$f^*(q^*) = \langle q^*, (\nabla f)^{-1}(q^*) \rangle - f((\nabla f)^{-1}(q^*)), \tag{27}$$

which is the *Legendre transform* of f [12, p. 35].

Remark 4. We have derived the Legendre transform and certain of its properties from a differentiable viewpoint, because the corresponding functions defined by the moment problem are in fact infinitely differentiable. In contrast, the trend in modern optimization theory is to assume as little differentiability as possible. For example, if f is a strictly convex C^1 function, then F is a continuous injection defined on U and is therefore an open mapping by Brouwer's Theorem on Invariance of Domain. Thus it is a *homeomorphm* between U and V. Following [14], one can define the conjugate function via (26) and verify that (27) holds. In particular, the inverse of F is given by a gradient. Far deeper is the situation when f maps U to an open convex set W, and one also wants to show the $V = W$. Such a global inverse function theorem for strictly convex C^1 functions f is given in the beautiful Theorem 26.5 in [14], under the additional assumption that f is a convex function of Legendre type. Returning to the case of smooth F, a global inverse function theorem can be proved under the condition that F is proper, in which case F is a *diffeomorphism.*

5 The main theorem

Applying the path integration methods of the previous section to the function F in Theorem 1, we obtain the strictly concave C^∞ function $f : \mathfrak{P}_+ \to \mathbb{R}$ taking the values

$$f(q) = \int_{\mathcal{I}} \Psi(t) \log \mathrm{Re}\{q(t)\} dt. \tag{28}$$

The function f can be extended to the closure of $\mathfrak{P}_+$ as an extended real-valued function. In particular, F is a diffeomorphism between $\mathfrak{P}_+$ and its open image in $\mathfrak{C}_+$.

In this setting,

$$\mathbb{J}_\Psi(\hat{q}(c)) = f^*(c), \tag{29}$$

where $\hat{q}(c)$ is the minimizer of (9) expressed as a function of c. According to Remark 2, f^* is strictly concave on any convex subset of $F(\mathfrak{P}_+)$, since f is. (See also [14, page 308] for a discussion about properties of the conjugate function f^* in the concave setting.) We also note that

$$\mathbb{I}_\Psi(\hat{\Phi}(c)) = 1 - \mathbb{J}_\Psi(\hat{q}(c)), \quad \text{for all } c \in \mathfrak{C}_+ \tag{30}$$

by Theorem 2, and hence, in view of (29), the function $\varphi : \mathfrak{C}_+ \to \mathbb{R}$, defined in (14), is given by

$$\varphi(c) = 1 - f^*(c), \tag{31}$$

and consequently φ is a strictly convex function on any convex subset of $F(\mathfrak{P}_+)$. We are now prepared to prove our main result.

Theorem 4. *The function F defined in Theorem 1 is a diffeomorphism between $\mathfrak{P}_+$ and $\mathfrak{C}_+$. Moreover, the value function φ is a convex function of Legendre type on $\mathfrak{C}_+$.*

Proof. Since the image of F is an open subset in the convex set $\mathfrak{C}_+$, it suffices to prove that it is also closed. To show this, we show that F is proper, i.e. that for any compact subset K of $\mathfrak{C}_+$, the inverse image $F^{-1}(K)$ is compact in $\mathfrak{P}_+$. This will follow from the fact that F is infinite on the boundary of $\mathfrak{P}_+$, which in turn follows from the following calculation:

$$\frac{\partial f}{\partial q_k} = \int_{\mathcal{I}} \alpha_k(t) \frac{\Psi(t)}{\mathrm{Re}\{q(t)\}} dt, \quad k = 0, 1, \dots, n. \tag{32}$$

Now, $t \mapsto \mathrm{Re}\{q(t)\}$ is a smooth, nonnegative function on $\mathcal{I}$. As q tends to the boundary of $\mathfrak{P}_+$, this function attains a zero on the interval, and hence, since $\alpha_0 = 1$ and q is C^2, the integral (32) is divergent at least for $k = 0$. Therefore, F is a diffeomorphism between $\mathfrak{P}_+$ and $\mathfrak{C}_+$.

We have already seen that φ is a strictly convex function. Therefore it just remains to show that f is essentially smooth. Clearly, $\mathfrak{P}_+$ is nonempty and f is differentiable throughout $\mathfrak{P}_+$, so conditions 1 and 2 in Definition 1 are satisfied. On the other hand, condition 3 is equivalent to properness of F, which we have already established above.

All that remains to be proven are the identities in Theorem 3. Recalling that the function $F : \mathfrak{P}_+ \to \mathfrak{C}_+$ in Theorem 1 is given by

$$F(q) = \nabla f(q), \tag{33}$$

the map $\nabla\varphi : \mathfrak{C}_+ \to \mathfrak{P}_+$ is the inverse of the diffeomorphism $-F$, i.e.,

$$\nabla\varphi = -F^{-1}. \tag{34}$$

Therefore, $\nabla\varphi$ sends c to $-\hat{q}(c)$, which establishes (15). To prove (16), observe that the Hessian

$$\frac{\partial^2\varphi}{\partial c^2} = -\frac{\partial\hat{q}}{\partial c}$$

but, since $F(\hat{q}) = c$, this is the inverse of

$$-\left.\frac{\partial F}{\partial q}\right|_{q=\hat{q}},$$

which, in view of (12), is precisely (16). Clearly, the strictly convex function φ has a unique minimum in the compact set $\mathcal{C}_0$. The minimizing point $\hat{c}$ belongs to the interior of $\mathcal{C}_0$ only if $\langle\nabla\varphi(\hat{c}), h\rangle = 0$ for all $h \in T_{\hat{c}}\mathcal{C}_+$, in which case we must have $\hat{q}(\hat{c}) = q_0 = 1$ by (15). This concludes the proof of Theorem 3.

Remark 5. As discussed in Remark 4, one can also deduce this theorem from Theorem 26.5 in Rockafeller [14], which would imply that F is a homeomorphism for a C^1 strictly convex function f. That F is a diffeomorphism for a C^2 function f would then follow from the Inverse Function Theorem. An alternative route, as indicated in Remark 4 could be based on Brouwer's Theorem on Invariance of Domain to prove that F is a homeomorphism. Either proof would of course entail the use of substantial additional machinery not needed in the smooth case. Indeed, this motivated us to develop the self-contained derivation of the Legendre transform and the subsequent proof presented here.

Acknowledgment

We would like to thank J. W. Helton for inquiring about the relationship between our earlier differential geometric proof of Theorem 1 and the Legendre transform, which led to the development of Sections 4 and 5.

References

1. N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*, Hafner Publishing, New York, 1965.
2. C. I. Byrnes and A. Lindquist, *Interior point solutions of variational problems and global inverse function theorems*, submitted for publication.
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: Festschrift in Honor of Hidenori Kimura on the Occasion of his 60th Birthday", Koichi Hashimoto, Yasuaki Oishi, and Yutaka Yamamoto, Editors, Birkhauser Boston, December 2002.

4. C. I. Byrnes, S.V. Gusev, and A. Lindquist, *A convex optimization approach to the rational covariance extension problem,* SIAM J. Control and Optimization **37** (1999), 211–229.
5. C. I. Byrnes, T. T. Georgiou and A. Lindquist, *A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint*, IEEE Trans. Automatic Control **AC-46** (2001), 822–839.
6. 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 Processing **SP-49** (Nov. 2000), 3189–3205.
7. C. I. Byrnes, S.V. Gusev, and A. Lindquist, *From finite covariance windows to modeling filters: A convex optimization approach,* SIAM Review **43** (Dec. 2001), 645–675.
8. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, 1991.
9. I. Csiszár, *I-divergence geometry of probability distributions and minimization problems,* Ann. Probab. **3** (1975), 146–158.
10. T.T. Georgiou and A. Lindquist, *Kullback-Leibler approximation of spectral density functions*, IEEE Trans. on Information Theory, to be published.
11. I. J. Good, *Maximum entropy for hypothesis formulation, especially for multidimentional contingency tables,* Annals Math. Stat. **34** (1963), 911–934.
12. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II*, Springer-Verlag, 1991.
13. S. Kullback, *Information Theory and Statistics*, 2nd edition, New York: Dover Books, 1968 (1st ed. New York: John Wiley, 1959).
14. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
15. C. R. Shankwitz and T T. Georgiou, *On the maximum entopy method for interval covariance sequences*, IEEE Trans. Acoustics, Speech and Signal processing **38** (October 1990), 1815–1817.