On Theorem 6 in “Relative Entropy and the Multivariable Multidimensional Moment Problem”

[Mar 2006 1052-1066]

Bin Zhu

Abstract—Matrix-valued covariance extension and multivariable spectral estimation are formulated as generalized moment problems in the “THREE” approach and its extensions. Under this context, we discuss Theorem 6 in [1] concerning the bijectivity of a moment map defined over a parametric family of spectral densities. In particular, we provide a counterexample in which the moment map under consideration is shown to have a critical point, namely a point at which the Jacobian loses rank. Then with standard techniques in bifurcation theory, we conclude further that the computed critical point is a bifurcation point, which means that the moment map is not injective.

Index Terms—Multivariate spectral estimation, parametrization of rational spectra, generalized moment problem, singular Jacobian, bifurcation point.

I. INTRODUCTION

MULTIVARIATE spectral estimation is an important and challenging problem in the fields of system identification, modeling, and signal processing. The “THREE” framework for spectral estimation first appeared in the seminal paper [2] by Byrnes, Georgiou, and Lindquist, which can be seen as a generalization of previous works on rational covariance extension and Nevanlinna-Pick interpolation; cf. [3]–[12] and references therein. Since then it has been significantly developed and extended to the multivariate case. We mention an incomplete list of contributions [13]–[18] in the scalar case, and [1], [19]–[31] for the multivariate counterpart. In that framework, the steady-state covariance matrix of the output process of a rational filter is used as data for the reconstruction of the input spectrum, which naturally admits a formulation as a generalized moment problem. Due to the typical ill-posedness of moment problems [32], [33], entropy-like functionals are then exploited as optimization criteria to promote uniqueness of the solution. More specifically, one tries to find the input spectrum consistent with the output covariance matrix that maximize some entropy or minimize some distance index to an a priori spectral density. A key feature of the approach is that parameter tuning is allowed in order to achieve high resolution in specified frequency bands.

Different choices of cost functionals lead to different forms of solutions, especially in the multivariate case; cf. [1], [19], [24]–[27]. Among them [1] is an important work utilizing the following relative entropy as the optimization criterion

$$S(\Phi|\Psi) = \int_{\mathbb{T}} \text{tr} \left[ \Phi (\log \Phi - \log \Psi) \right]$$

which in turn, draws inspiration from quantum mechanics. Here $\Psi$ is the known prior and $\mathbb{T}$ stands for the unit circle. Minimization of $S(\Phi|\Psi)$ with respect to $\Phi$ subject to the generalized moment constraint can be worked out explicitly leading to an exponential-type spectral density. Such a solution can also be recovered as a limit case of a family of solutions based on the multivariate Beta divergence discussed in [26]. Difficulty arises in the other direction, namely, minimization of $S(\Psi|\cdot)$ with respect to the second argument. As reported in [1], variational analysis and duality reasoning hit an obstruction in the middle because the functional dependence of the optimal primal on the dual variable cannot be described in a closed form; see also [21]. As a response to this difficulty, Theorem 6 of [1] suggests to “forgo an explicit form for the entropy functional and start instead with a computable Jacobian”. In other words, a parametric form of the spectral density has been proposed, which possibly does not correspond to any cost functional. Although the statement of that theorem looks rather exciting, it is extremely nontrivial and its validity remains elusive as a rigorous proof is absent. In this note, we are motivated to address this issue. We shall only consider the first half of [1] Theorem 6 concerning rational solutions to the spectral estimation problem.

The continuation argument is used extensively in the proofs of [1] which follows the previous work [14] in the scalar case by the same author. As will be reviewed later in Section IV in order for the argument to be effective, the Jacobian of the parametric moment map is required to vanish nowhere in the feasible set. In this work, we show through a numerical example that this requirement is not met in general by the moment map in question, contrary to what is claimed in [1]. Furthermore, a critical point of the moment map is computed and demonstrated to be a bifurcation point. In consequence, the parametric solution to the spectral estimation problem considered in [1] Section IV] is generally not unique.

This note is organized as follows. In Section II we review the parametric form of the moment map introduced in [1] that will be the central object of investigation in this work. We give a numerical example in Section III where a critical point of the moment map is detected and computed. In Section IV we apply a part of the bifurcation theory and carry out some
further computation which allows us to conclude that the afore
obtained critical point is in fact a bifurcation point. Finally,
we make some remarks on an alternative parametrization of
rational spectral densities.

II. PROBLEM REVIEW

One of the problems considered in [1] is about finding
a matrix spectral density function in a particular parametric
family that satisfies a (generalized) moment constraint. In
order to restate one of the main results of that paper, it is
necessary to introduce some notations first:
• \( G(z) = (zI - A)^{-1}B \) is a rational filter defined by
the matrix pair \((A, B)\) such that \( A \in \mathbb{C}^{n \times n}\)
is a stability matrix in the discrete-time sense, \( B \in \mathbb{C}^{n \times m}\)
has full column rank, and \((A, B)\) is reachable.
• \( \mathcal{H}_n \) is the vector space of \( n \times n \) Hermitian matrices
over the reals; \( \mathcal{H}_{n^+} \) contains positive definite matrices.
• \( C(\mathbb{T}; \mathcal{H}_m) \) is the vector space of continuous \( \mathcal{H}_m\)-valued
function on the unit circle \( \mathbb{T} := \{ z \in \mathbb{C} : |z| = 1 \} \).
•

\[
\Gamma : \Phi \mapsto \int G\Phi G^* \tag{1}
\]
is an operator from \( C(\mathbb{T}; \mathcal{H}_m) \) to \( \mathcal{H}_n \), where the integral \( \int F \) is a shorthand for \( \int_0^1 F(e^{i\theta}) d\theta \). The range of \( \Gamma \), denoted by \( \text{Range } \Gamma \), is a subspace of \( \mathcal{H}_n \). The symbol used in [1] for the same operator is \( L \).
• \( \mathcal{L}_+ := \{ A \in \mathcal{H}_n : G^*(z)\Lambda G(z) > 0, \forall z \in \mathbb{T} \} \).
• \( \Psi \) is a bounded and coercive \( m \times m \) spectral density
function, that is, there exist constants \( \mu, M \) such that
\( \mu I \leq \Psi(e^{j\theta}) \leq MI \) for all \( \theta \in (-\pi, \pi] \). It admits a
(unique) left outer factor \( W_\Psi \), namely, \( \Psi = W_\Psi W_{\Psi}^* \).

The notations used in [1] for \( \Psi \) and its factor are \( \sigma \) and
\( \sigma^{1/2} \), respectively.
•

\[
h : \Lambda \mapsto \int GW_\Psi (G^* \Lambda G)^{-1}W_{\Psi}^* G^* \tag{2}
\]
is a map from \( \mathcal{L}_+ := \mathcal{L}_+ \cap \text{Range } \Gamma \) to \( \text{Range } \Gamma := \mathcal{H}_{n^+} \cap \text{Range } \Gamma \). The domain and codomain of the map are denoted with \( \text{Ker}(C) \) and \( \text{int}(C) \) in [1], respectively. Moreover, the argument \( \Lambda \) is lower cased in [1].

Theorem 6 of [1] states that the map \( h \) is a bijection given
any bounded and coercive prior \( \Psi \). In other words, given
any positive definite matrix \( \Sigma \in \text{Range } \Gamma \), there exists a unique
parameter \( \Lambda \in \mathcal{L}_+ \) such that the spectral density
\[
\Phi = W_{\Psi}(G^* \Lambda G)^{-1}W_{\Psi}^* \tag{3}
\]
solves the generalized moment equation \( \Gamma(\Phi) = \Sigma \). A key argument in that paper is that the Jacobian \( \nabla h(\Lambda) : \text{Range } \Gamma \to \text{Range } \Gamma \) is invertible for any \( \Lambda \in \mathcal{L}_+ \). We will provide a two-
dimensional \((m = 2)\) numerical counterexample in the next
section to this argument showing that the Jacobian of \( h \) can be singular at one point.

III. SINGULAR JACOBIAN OF THE MOMENT MAP

The Jacobian of the moment map \( h \), i.e., its Fréchet deriva-
tive, is a linear operator from \( \text{Range } \Gamma \) to itself:
\[
\nabla h(\Lambda) : \delta \Lambda \mapsto - \int GW_{\Psi}(\Lambda)^{-1} G^*(\delta \Lambda) \Gamma(\Lambda)^{-1}W_{\Psi} G^* \tag{4}
\]
where \( \Gamma^* : X \to G^* X G \) is the adjoint operator of \( \Gamma \) in [1] from \( \mathcal{H}_n \) to \( C(\mathbb{T}; \mathcal{H}_m) \), and \( \Gamma^*(\Lambda)^{-1} \) is understood as
\( (G^* \Lambda G)^{-1} \).

The claim that \( \nabla h(\Lambda) \) vanishes nowhere in \( \mathcal{L}_+ \) is true in
the special case when the prior \( \Psi = \psi I \), namely, a scalar
spectral density function times the identity matrix. Details
can be found in [1] itself; see also [23, 41]. An important observation is that the Jacobian in that case is a self-adjoint
operator, and in fact, it is equal to the negative Hessian of a
certain cost function. Therefore, the reasoning of nonvanishing
Jacobian is built upon the definiteness of the quadratic form
\( (\delta \Lambda, \nabla h(\Lambda)(\delta \Lambda)) \), where the standard inner product over \( \mathcal{H}_n \) is defined as \( \langle A, B \rangle := \text{tr}(AB) \). Such reasoning fails in
general when \( \Psi \) is arbitrarily (but fixed) matrix-valued because
the self-adjoint property is lost. One can simply verify that
the adjoint operator \( \nabla h(\Lambda)^* : \text{Range } \Gamma \to \text{Range } \Gamma \) of the
Jacobian \( h \) is given by
\[
\delta \Lambda \mapsto - \int GT(\Lambda)^{-1} \Gamma^* (\delta \Lambda) \Gamma(\Lambda)^{-1} G^* \tag{5}
\]
which is different from \( \nabla h(\Lambda) \).

In the sequel, we want to evaluate numerically the Jacobian
determinant. Before that, we will have to build a matrix
representation of the linear operator \( \nabla h(\Lambda) \).

A. Matrix representation of the Jacobian

The Jacobian \( h \) is a linear map from a finite dimensional
vector space to itself. It admits a matrix representation if we
fix an orthonormal basis of \( \text{Range } \Gamma \), say \( \{A_k\}_{k=1}^M \),
with \( M = m(2n - m) \) the complex case (cf. [34] Proposition
3.1) for the dimension). More precisely, the \((j, k)\) element of
the real \( M \times M \) Jacobian matrix \( J_h(\Lambda) \) is
\[
\langle A_j, \nabla h(\Lambda)(A_k) \rangle \tag{5}
\]
The domain of the map \( h \), namely the set \( \mathcal{L}_+ \), is convex,
which is in particular path-connected. We have the next simple
proposition.

Proposition 1. Consider a \( C^1 \) map \( f : D \subset \mathbb{R}^n \to \mathbb{R}^n \) such
that \( D \) is path-connected. If its Jacobian \( \nabla f : D \to \mathbb{R}^{n \times n} \) is
everywhere nonsingular, then its determinant \( \det \nabla f(x) \) does
not change sign over \( D \).

Proof. Suppose the contrary, i.e., there exist two points
\( x_1, x_2 \in D \) such that \( \det \nabla f(x_1) > 0 \) and \( \det \nabla f(x_2) < 0 \). By the assumption of path-connectedness, there exists a
continuous function \( p : [0, 1] \to D \) such that \( p(0) = x_1 \)
and \( p(1) = x_2 \). Since \( f \) is \( C^1 \), the real-valued function
\( \det \nabla f(p(t)) \) is continuous. By the intermediate value theorem
it must be zero for some \( t \in (0, 1) \).

Therefore, if a sign change of the Jacobian determinant
is detected, the Jacobian of the map under consideration
cannot be everywhere nonsingular. This is the idea behind our numerical example.

B. A numerical example

Here we consider the problem of matrix covariance extension of dimension \( m = 2 \) with maximal covariance lag \( p = 1 \), the (probably) simplest nontrivial case. We have \( n = m(p + 1) = 4 \). The matrix pair \((A, B)\) of the filter bank \( G(z) \) is given by

\[
A = \begin{bmatrix} 0 & I_2 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ I_2 \end{bmatrix}, \quad \text{with} \quad G(z) = \begin{bmatrix} z^{-2}I_2 \\ z^{-1}I_2 \end{bmatrix}.
\]

Let us work in the real case, in which \( \text{Range} \Gamma \) is the \( M = 7 \)-dimensional vector space of symmetric block-Toeplitz matrices of the form

\[
\begin{bmatrix}
\Lambda_0 & \Lambda^T_1 \\
\Lambda_1 & \Lambda_0
\end{bmatrix},
\]

where \( \Lambda_0, \Lambda_1 \) are \( 2 \times 2 \) blocks. An orthogonal but unnormalized basis of \( \text{Range} \Gamma \) can be determined from matrix pairs

\[
(\Lambda_0, \Lambda_1) \in \{0\} \times \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \times \{0\},
\]

(6)

where the bold symbol \( \mathbf{0} \) denotes the \( 2 \times 2 \) zero matrix. Normalization of the basis matrices is necessary to compute the quantity \( \mathbf{5} \) correctly.

The prior is taken as \( \Psi = KGG^*K^* \), a matrix Laurent polynomial, for

\[
K = \begin{bmatrix}
-0.22 & -1.23 & 2.22 & 0 \\
-1.11 & 0.96 & 1.14 & 2.49
\end{bmatrix}.
\]

The polynomial \( zKG \) is Schur, with determinantal roots 0.5868, -0.3558, and thus the outer factor \( W_\Psi = zKG \) in this example.

The argument \( \Lambda \) lives in the open set \( \mathcal{L}_1^{+} \). In practice, it is better to start with a factor of the form \( zCG \) with \( C \in \mathbb{C}^{m \times n} \). Then we can form the function

\[
G^* \Lambda G := G^* C^* C G.
\]

(7)

Notice that if we assign the elements of \( C \) with Gaussian or uniformly distributed random numbers, it is highly unlikely that the polynomial \( \det zCG \) has roots on the unit circle. From (7) we have the relation that \( \Lambda \) is equal to the projection of \( C^*C \) onto the subspace \( \text{Range} \Gamma \). Details of the spectral factorization (7) can be found in [23, 24, 31].

We have picked two \( C \) matrices with corresponding \( \Lambda \) matrices and the determinantal roots of \( zCG \) reported below:

\[
C^{(0)} = \begin{bmatrix}
-1.08 & -0.57 & 2.45 & 0 \\
0.84 & -0.08 & 1.01 & 0.78
\end{bmatrix},
\]

(8)

where \( \text{det} zCG \) at 0.1211 ± 0.5302i (modulus 0.5438).

\[
C^{(1)} = \begin{bmatrix}
0.63 & 0.67 & 1.45 & 0 \\
1.68 & -0.61 & 1.04 & 2
\end{bmatrix}
\]

(9)

with the roots of \( \text{det} zC^{(0)}G \) at 0.1211 ± 0.5302i (modulus 0.5438).

\[
C^{(1)} = \begin{bmatrix}
0.63 & 0.67 & 1.45 & 0 \\
1.68 & -0.61 & 1.04 & 2
\end{bmatrix}
\]

(10)

corresponds to the blocks of \( \Lambda^{(1)} \)

\[
\Lambda_0^{(1)} = \begin{bmatrix}
3.2017 & 0.7387 \\
0.7387 & 2.4105
\end{bmatrix}, \quad \Lambda_1^{(1)} = \begin{bmatrix}
2.6607 & 0.3371 \\
3.3600 & -1.2200
\end{bmatrix}
\]

(11)

with the roots of \( \text{det} zC^{(1)}G \) at 0.7791, -0.6683.

The integral in (4) is approximated with the Riemann sum in Matlab:

\[
\int F(\theta) \approx \frac{\Delta \theta}{2\pi} \sum_k F(\theta_k),
\]

(12)

where \( \{\theta_k\} \) are equidistant points on the interval \( (-\pi, \pi] \) and the “step length” \( \Delta \theta = 10^{-4} \). With the normalized basis obtained from (6), the Jacobian matrix can be computed explicitly as in (5) and its determinant can be evaluated. We have the numerical result \( \det J_k(\Lambda^{(k)}) = 10.6871, -326.6439 \) for \( k = 0, 1 \), respectively.

Computation of the above example has also been implemented in Mathematica in order to evaluate the integrals symbolically given the numerical values of \( \Lambda \). The result is consistent with the numerical computation in Matlab, i.e., a sign change of the Jacobian determinant has been detected.

Further, the critical point \( \Lambda^C \) can be computed using the bisection method on the real-valued function \( \det J_k(\Lambda^{(t)}) \) where

\[
\Lambda^{(t)} = (1 - t)\Lambda^{(0)} + t\Lambda^{(1)}, \quad t \in [0, 1]
\]

(13)

is the line segment between \( \Lambda^{(k)}, \ k = 0, 1 \). We have the blocks

\[
\Lambda_0^C = \begin{bmatrix}
4.3901 & 0.6713 \\
0.6713 & 0.5589
\end{bmatrix}, \quad \Lambda_1^C = \begin{bmatrix}
-1.5930 & -1.3940 \\
0.7793 & -0.1155
\end{bmatrix},
\]

(14)

with the corresponding \( t^C = 0.0459, \det J_k(\Lambda^C) = -5.4964 \times 10^{-14} \), and the two smallest singular values of \( J_k(\Lambda^C) \) are 1.1053 \times 10^{-16} and 0.0573. Hence the Jacobian matrix of \( h \) loses exactly rank 1 at \( \Lambda^C \).

IV. CHARACTERIZATION OF THE CRITICAL POINT

The quest for nowhere vanishing Jacobian is motivated by the use of continuation methods to solve the nonlinear equation \( h(\Lambda) = \pi \) for the parameter \( \Lambda \). The idea is briefly reviewed in the next proposition when the map under consideration is a diffeomorphism; cf. [33] for more general settings.

Proposition 2. Assume for simplicity that \( D, E \) are open and convex subsets of \( \mathbb{R}^n \). Let \( f : D \to E \) be a \( C^1 \) diffeomorphism. Then for \( y \in E \), the solution \( x = f^{-1}(y) \) can be found by solving the initial value problem

\[
\begin{cases}
\dot{x}(t) = [\nabla f(x(t))]^{-1} (y - y_0) \\
x(0) = x_0
\end{cases}
\]

(15)

and evaluating \( x := x(1) \). The initial value \( x_0 \in D \) is arbitrary and \( y_0 := f(x_0) \).

Proof. By the assumption of convexity, the line segment

\[
p(t) = ty_0 + (1 - t)x_0, \quad t \in [0, 1]
\]

(16)
is inside $E$. It is easy to verify that the solution curve $x(t) := f^{-1}(p(t))$ satisfy the IVP $[9]$. In fact, the differential equation comes from differentiating the two sides of $f(x(t)) = p(t)$ w.r.t. $t$ and inverting the Jacobian $\nabla f(x(t))$. Due to the assumption that $f$ is a diffeomorphism, the solution curve $x(t)$ is indeed continuously differentiable and the Jacobian of $f$ is everywhere invertible in $D$.

The precise terminal point $x(1)$ can be obtained using a predictor-corrector algorithm $[35, 36]$. If one is satisfied enough with an approximate solution, then a general-purpose ODE solver can be used to numerically integrate $[9]$. Of course, the map $f$ in the above proposition being a diffeomorphism is a sufficient condition for the continuation method to return a unique solution. This is indeed the case for our $h$ map when the prior takes the special form $\Psi = \psi I$ as mentioned previously; cf. $[31]$. However, in the presence of a singular Jacobian, it can happen that the solution curve to the IVP branches out at a critical point, and several terminal points exist. On the other hand, a numerical ODE solver diverges in that case because the norm of the derivative tends to infinity near the critical point.

Next we shall demonstrate numerically that the critical point computed in Section $[11]$ is a bifurcation point. To this end, it is customary to define the augmented map

$$\mathcal{H}(\Lambda, t) := h(\Lambda) - p(t)$$

from $\mathbb{R}^L \times [0, 1] \to \text{Range } \Gamma$, where $p(t) := h(\Lambda(t))$ is a smooth curve in $\text{Range } \Gamma$ with $\Lambda(t)$ in $[8]$. Under this convention, the curve $(\Lambda(t), t)$ parametrized by $t$ is in the zero set $H^{-1}(0)$. When a basis of $\text{Range } \Gamma$ is fixed as in the previous section, the map $\mathcal{H}$ can be identified as a function $H$ from a subset of $\mathbb{R}^{M+1}$ to $\mathbb{R}^M$, whose coordinates have the expression

$$H_j : (\Lambda, t) \mapsto \langle \Lambda_j, H(\Lambda, t) \rangle, \quad j = 1, \ldots, M,$$

where $\Lambda = \sum_k x_k \Lambda_k$ with $x \in \mathbb{R}^M$ the coordinate vector. Explicit calls of the coordinate vector $x$ will be avoided subsequently in order to ease the notation.

The matrix representation of the augmented Jacobian $J_H \in \mathbb{R}^{M \times (M+1)}$ can be described in terms of the following vector with each entry in $\text{Range } \Gamma$:

$$\nabla H(\Lambda, t) = \left[ \nabla h(\Lambda)(\Lambda_1) \cdots \nabla h(\Lambda)(\Lambda_M) \right] - \dot{p}(t),$$

where $\dot{p}(t) = \nabla h(\Lambda(t))(\Lambda(1) - \Lambda(0))$. Then the $(j,k)$ element of the augmented Jacobian

$$[J_H(\Lambda, t)]_{jk} = \langle \Lambda_j, \nabla H(\Lambda, t) \rangle_k.$$ 

Notice that the last column of $J_H(\Lambda, t)$ does not increase the rank due to the relation $\Lambda(t^*) = \Lambda_c$. Hence we have

$$\text{rank } J_H(\Lambda_c, t^*) = M - 1, \quad \text{dim } \text{Ker } J_H(\Lambda_c, t^*) = 2.$$

Let us introduce the Lyapunov-Schmidt reduction in our finite dimensional context:

$$\mathbb{R}^{M+1} = D_1 \oplus D_2, \quad \mathbb{R}^M = E_1 \oplus E_2,$$

where

$$D_1 := \ker J_H(\Lambda_c, t^*), \quad D_2 := D_1^\perp, \quad E_2 := \text{Range } J_H(\Lambda_c, t^*), \quad E_1 := E_2^\perp.$$ 

The above subspaces can be made more precise by performing SVD to the Jacobian matrix of $H$ at $(\Lambda_c, t^*)$, namely

$$J_H(\Lambda_c, t^*) = U \Sigma V^T$$

where $\Sigma$ is the (square) diagonal matrix containing all the nonzero singular values, $\sigma, \nu$ are columns of the orthogonal matrices $U, V$, respectively, and the notation $u_{jk}$ denotes the matrix obtained by putting together the columns $u_j, u_{j+1}, \ldots, u_k$. It is then elementary to verify that

$$D_1 = \text{Range } V_2, \quad D_2 = \text{Range } V_1, \quad E_2 = \text{Range } U_1, \quad E_1 = \text{Range } U_2.$$ 

We can then partition $H$ w.r.t. the new bases determined by the singular vectors. Specifically, let us define

$$\tilde{H}(y) = \begin{bmatrix} \tilde{H}_1(y_1, y_2) \\ \tilde{H}_2(y_1, y_2) \end{bmatrix} := \begin{bmatrix} U_2^T \\ U_1 \end{bmatrix} H(V_2 y_1 + V_1 y_2),$$

where $y = (y_1, y_2) \in \mathbb{R}^2 \times \mathbb{R}^{M-1}$ are coordinates of the argument vector $(\Lambda, t)$ in $[10]$ under the new basis. The Jacobian of $H$ is computed as

$$\nabla \tilde{H}(y) = \left[ \nabla \tilde{H}_1(y_1, y_2) \nabla \tilde{H}_2(y_1, y_2) \right] = \begin{bmatrix} U_2^T \\ U_1 \end{bmatrix} \nabla H(\Sigma V_2 y_1 + \Sigma V_1 y_2) [V_2, V_1],$$

where $\nabla \tilde{H}_j$ denotes the Jacobian of $\tilde{H}_j$ w.r.t. the variable $y_k$. It is then straightforward to check that

$$\nabla \tilde{H}(y^*) = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $y^*$ is the coordinate of $(\Lambda_c, t^*)$ and $\Sigma_{M-1}$ nonsingular. Since we have $\tilde{H}_2(y_1^*, y_2^*) = 0$, the implicit function theorem can be applied to assert that locally around $y^*$

$$\tilde{H}_2(y_1, y_2) = 0 \iff y_2 = \varphi(y_1)$$

for some smooth function $\varphi$. Substituting $y_2$ with the above local functional dependence on $y_1$ into the equation $\tilde{H}_1(y_1, y_2) = 0$, we obtain that equivalently,

$$b(y_1) := \tilde{H}_1(y_1, \varphi(y_1)) = 0,$$

which is called the bifurcation equation at the critical point $y^*$ of $H$. Notice that $b$ is a real-valued function defined on some subset of $\mathbb{R}^2$. According to $[35]$ Definition 8.1.11, if the Hessian matrix $\nabla^2 b(y_1^*)$ has two eigenvalues of distinct signs, then $y^*$ is a simple bifurcation point of the equation $\tilde{H}(y) = 0$.

Following the derivation in $[35$ pp. 77-78], we have the equality

$$\nabla^2 b(y_1^*) = \nabla^2 \tilde{H}_1(y_1^*).$$
We now need a computable expression for the Hessian matrix. Its operator form is easily obtained
\[
\begin{align*}
\nabla^2 H_1(y) : (\delta y_{11}, \delta y_{12}) & \mapsto U_2^T H(V_{2y_1} + V_{1y_2})(2V_{2\delta y_{11}, 1} + 2V_{2\delta y_{12}}), \\
\end{align*}
\]
as a bilinear map from \(\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}\), whose matrix representation follows immediately
\[
\nabla^2 H_1(y) = V_2^T \left[ \sum_j u_j M \nabla^2 H_j(V_{2y_1} + V_{1y_2}) \right] V_2, \quad (12)
\]
where \(U_2 \equiv u_M\) is the last left singular vector in (11), and \(\nabla^2 H_j\) is the Hessian of the component function in (10).

Therefore, computation is ultimately reduced to evaluating the 3-d array of second-order partials \(\nabla^2 H\) under the standard (to be normalized) basis introduced in (6). Define the symmetric matrix with Range \(\Gamma\)-valued entries \(\nabla^2 H(\Lambda, t) := \nabla^2 h(\Lambda) = \frac{\partial^2 h(\Lambda)}{\partial \Lambda_1 \partial \Lambda_2} = \int F + F^*\)
is the second-order differential of \(h\) with
\[
\begin{align*}
F := GW^* \frac{\partial^* (\Lambda)}{\partial \Lambda_1} \frac{\partial^* (\Lambda)}{\partial \Lambda_2} G + \frac{\partial^* (\Lambda)}{\partial \Lambda_1} \frac{\partial^* (\Lambda)}{\partial \Lambda_2} G \quad \text{and} \quad \tilde{p}(t) = \frac{\partial^2 h(\Lambda)}{\partial \Lambda_1 \partial \Lambda_2},
\end{align*}
\]
The Hessian matrix of the component function results from taking element-wise inner product with (13), i.e.,
\[
[\nabla^2 H_j(\Lambda, t)]_{k,\ell} = \langle \Lambda_j, \nabla^2 H(\Lambda, t) \rangle_{k,\ell}, \quad k, \ell = 1, \ldots, \text{M + 1}.
\]
Continuing our numerical example in the previous section, the Hessian matrix \(\nabla^2 b(y_f^*)\) is computed according to the formula (12) and its two eigenvalues are \(-0.3226, 0.0239\). Therefore, we confirm that \(y_f^*\), or equivalently \((\Lambda^*, t^*)\), is a bifurcation point. Following the very definition of a bifurcation point [35, pp. 76], the original map \(h\) in (2) is not injective.

Remark 1. The sole purpose of the computation above is to show that the Hessian matrix \(\nabla^2 b(y_f^*)\) is nonsingular, which according to [35, pp. 78] is generic. In this case, the Hessian cannot have two eigenvalues of the same sign, since otherwise \((\Lambda^*, t^*)\) would be an isolated zero point of \(H\) which cannot be reached through curve tracing. This is a consequence of the celebrated theorem of Morse [35, Lemma 8.1.10].

V. CONCLUDING REMARKS
Although only nonvanishing Jacobian is emphasized in [11], properness is another important property of the moment map, as it is closely related to the question of surjectivity; cf. [30]. The argument on properness has been made implicitly when the prior is taken to be \(\Psi = I\), as can be seen in the second column of [11, pp. 1060], the part proving that the solution to the IVP can be “continued” until \(t = 1\). However, in the general case of a matrix-valued prior, a proof of the \(h\) map being proper does not seem obvious.

It is also worth pointing out that the solution form (3) to the moment problem plays a major role in [37], where the factor of \(\Psi\) is taken as \(W_\Psi = I + KG\) for some \(K \in \mathbb{C}^{m \times n}\), which is certainly matrix-valued, i.e., not scalar times identity.

At last, we would like to mention an alternative parametrization of rational spectral densities discussed in [23], [30], [31], where the “denominator” \(G^*AG\) is factored instead of breaking the prior down into factors as in (3). The moment map becomes
\[
\tau : C \mapsto \int G(C)^{-1} \Psi(C) G^*, \quad (14)
\]
where the parameter \(\tau\) determines the unique right outer factor of \(G^*AG\) as indicated in (7). It has been shown in [30] that the map \(\tau\) is surjective. Moreover, the derivative of (14) can be written down explicitly [31], and a singular Jacobian has so far not been detected numerically, which suggests that there is still hope for uniqueness in this alternative parametrization.

ACKNOWLEDGMENT

The author would like to thank Dr. Giacomo Baggio for implementing the numerical example in Mathematica.

REFERENCES

[1] T. T. Georgiou, “Relative entropy and the multivariable multidimensional moment problem,” IEEE Trans. Inform. Theory, vol. 52, no. 3, pp. 1052–1066, 2006.
[2] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, “A new approach to spectral estimation: A tunable high-resolution spectral estimator,” IEEE Trans. Signal Process., vol. 48, no. 11, pp. 3189–3205, 2000.
[3] R. E. Kalman, “Realization of covariance sequences,” in Toeplitz Centennial. Springer, 1982, pp. 331–342.
[4] T. T. Georgiou, “Partial realization of covariance sequences,” Ph.D. dissertation, University of Florida, Gainesville, 1983.
[5] ——, “Realization of power spectra from partial covariance sequences,” IEEE Trans. Acoust. Speech Signal Process., vol. 35, no. 4, pp. 438–449, 1987.
[6] ——, “A topological approach to Nevanlinna–Pick interpolation,” SIAM J. Math. Anal., vol. 18, no. 5, pp. 1248–1260, 1987.
[7] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev, “A complete parameterization of all positive rational extensions of a covariance sequence,” IEEE Trans. Automat. Control, vol. 40, no. 11, pp. 1841–1857, 1995.
[8] C. I. Byrnes and A. Lindquist, “On the partial stochastic realization problem,” IEEE Trans. Automat. Control, vol. 42, no. 8, pp. 1049–1070, 1997.
[9] C. I. Byrnes, S. V. Gusev, and A. Lindquist, “A convex optimization approach to the rational covariance extension problem,” SIAM J. Control Optim., vol. 37, no. 1, pp. 211–229, 1998.
[10] T. T. Georgiou, “The interpolation problem with a degree constraint,” IEEE Trans. Automat. Control, vol. 44, no. 3, pp. 631–635, 1999.
[11] C. I. Byrnes, S. V. Gusev, and A. Lindquist, “From finite covariance windows to modeling filters: A convex optimization approach,” SIAM Rev., vol. 43, no. 4, pp. 645–675, 2001.
[12] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, “A generalized entropy criterion for Nevanlinna–Pick interpolation with degree constraint,” IEEE Trans. Automat. Control, vol. 46, no. 6, pp. 822–839, 2001.
[13] T. T. Georgiou and A. Lindquist, “Kullback–Leibler approximation of spectral density functions,” IEEE Trans. Inform. Theory, vol. 49, no. 11, pp. 2910–2917, 2003.
[14] T. T. Georgiou, “Solution of the general moment problem via a one-parameter imbedding,” IEEE Trans. Automat. Control, vol. 50, no. 6, pp. 811–826, 2005.

[15] M. Pavon and A. Ferrante, “On the Georgiou–Lindquist approach to constrained Kullback–Leibler approximation of spectral densities,” IEEE Trans. Automat. Control, vol. 51, no. 4, pp. 639–644, 2006.

[16] P. Fjørtoft and J. Karlsson, “Minimal Itakura-Saito distance and covariance interpolation,” in 47th IEEE Conference on Decision and Control (CDC 2008). IEEE, 2008, pp. 137–142.

[17] M. Zorzi, “Rational approximations of spectral densities based on the alpha divergence,” Math. Control Signals Systems, vol. 26, no. 2, pp. 259–278, 2014.

[18] A. Ferrante, F. Ramponi, and F. Ticozzi, “On the convergence of an efficient algorithm for Kullback–Leibler approximation of spectral densities,” IEEE Trans. Automat. Control, vol. 56, no. 3, pp. 506–515, 2011.

[19] T. T. Georgiou, “Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parametrization,” IEEE Trans. Automat. Control, vol. 47, no. 11, pp. 1811–1823, 2002.

[20] ———, “The structure of state covariances and its relation to the power spectrum of the input,” IEEE Trans. Automat. Control, vol. 47, no. 7, pp. 1056–1066, 2002.

[21] A. Ferrante, M. Pavon, and F. Ramponi, “Hellinger versus Kullback–Leibler multivariable spectrum approximation,” IEEE Trans. Automat. Control, vol. 53, no. 4, pp. 954–967, 2008.

[22] F. Ramponi, A. Ferrante, and M. Pavon, “A globally convergent matricial algorithm for multivariate spectral estimation,” IEEE Trans. Automat. Control, vol. 54, no. 10, pp. 2376–2388, 2009.

[23] A. Ferrante, M. Pavon, and M. Zorzi, “Application of a global inverse function theorem of Byrnes and Lindquist to a multivariable moment problem with complexity constraint,” in Three Decades of Progress in Control Sciences. Springer Berlin Heidelberg, 2010, pp. 153–167.

[24] E. Avventi, “Spectral moment problems: Generalizations, implementation and tuning,” Ph.D. dissertation, KTH Royal Institute of Technology, Stockholm, 2011.

[25] A. Ferrante, C. Masiero, and M. Pavon, “Time and spectral domain relative entropy: A new approach to multivariate spectral estimation,” IEEE Trans. Automat. Control, vol. 57, no. 10, pp. 2561–2575, 2012.

[26] M. Zorzi, “A new family of high-resolution multivariate spectral estimators,” IEEE Trans. Automat. Control, vol. 59, no. 4, pp. 892–904, 2014.

[27] ———, “Multivariate spectral estimation based on the concept of optimal prediction,” IEEE Trans. Automat. Control, vol. 60, no. 6, pp. 1647–1652, 2015.

[28] T. T. Georgiou and A. Lindquist, “Likelihood analysis of power spectra and generalized moment problems,” IEEE Trans. Automat. Control, vol. 62, no. 9, pp. 4580–4592, 2017.

[29] G. Baggio, “A global convergence analysis of the Pavon–Ferrante algorithm for spectral estimation,” to appear in IEEE Trans. Automat. Control, 2017.

[30] B. Zhu and G. Baggio, “On the existence of a solution to a spectral estimation problem à la Byrnes-Georgiou-Lindquist,” to appear in IEEE Trans. Automat. Control, arXiv e-print https://arxiv.org/abs/1709.09012, 2017.

[31] B. Zhu, “On a parametric spectral estimation problem,” accepted for presentation at the 18th IFAC Symposium on System Identification (SYSID 2018), arXiv e-print https://arxiv.org/abs/1712.07970, 2017.

[32] U. Grenander and G. Szegö, Toeplitz Forms and Their Applications, ser. California Monographs in Mathematical Sciences. University of California Press, 1958.

[33] M. G. Krein and A. A. Nudel’man, The Markov Moment Problem and Extremal Problems, ser. Translations of Mathematical Monographs. Providence, Rhode Island: American Mathematical Society, 1977, vol. 50.

[34] A. Ferrante, M. Pavon, and M. Zorzi, “A maximum entropy enhancement for a family of high-resolution spectral estimators,” IEEE Trans. Automat. Control, vol. 57, no. 2, pp. 318–329, 2012.

[35] E. L. Allgower and K. Georg, Numerical Continuation Methods: An Introduction, ser. Springer Series in Computational Mathematics. Springer-Verlag, 1990, vol. 13.

[36] B. Zhu, “On the well-posedness of a generalized moment problem and its numerical solution,” submitted to IEEE Trans. Automat. Control, arXiv e-print https://arxiv.org/abs/1802.09330, 2018.

[37] M. S. Takiar and T. T. Georgiou, “Analytic interpolation with a degree constraint for matrix-valued functions,” IEEE Trans. Automat. Control, vol. 55, no. 5, pp. 1075–1088, 2010.