Systems & Control Letters 59 (2010) 167–172

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

On the well-posedness of multivariate spectrum approximation and convergence of high-resolution spectral estimators[☆]

Federico Ramponi[a,*], Augusto Ferrante[b], Michele Pavon[c]

[a] Institut für Automatik, ETH Zürich, Physikstrasse 3, 8092 Zürich, Switzerland

[b] Dipartimento di Ingeneria dell'Informazione, Università di Padova, via Gradenigo 6/B, I-35131 Padova, Italy

[c] Dipartimento di Matematica Pura e Applicata, Università di Padova, via Trieste 63, 35131 Padova, Italy

ARTICLE INFO

Article history:
Received 12 March 2009
Received in revised form
29 December 2009
Accepted 11 January 2010
Available online 20 February 2010

Keywords:
Spectrum estimation
Spectrum approximation
Well-posedness
Consistency

ABSTRACT

In this paper, we establish the well-posedness of the generalized moment problems recently studied by Byrnes–Georgiou–Lindquist and coworkers, and by Ferrante–Pavon–Ramponi. We then apply these continuity results to prove the almost sure convergence of a sequence of high-resolution spectral estimators indexed by the sample size.

1. Introduction

Consider a linear, time invariant system

$$x(t+1) = Ax(t) + By(t), \quad A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, \tag{1}$$

with the transfer function

$$G(z) = (zI - A)^{-1}B, \tag{2}$$

where A is a stability matrix, B is full column rank, and (A, B) is a reachable pair. Suppose that the system is fed with a m-dimensional, zero-mean, wide-sense stationary process y having a spectrum Φ. The asymptotic state covariance Σ of the system (1) satisfies:

$$\Sigma = \int G\Phi G^*. \tag{3}$$

Here and in the following, $G^*(z) = G^\top(z^{-1})$, and integration takes place over the unit circle with respect to the normalized Lebesgue measure $\mathrm{d}\vartheta/2\pi$. Let $\mathcal{S}_+^{m\times m}(\mathbb{T})$ be the family of bounded, coercive, $\mathbb{C}^{m\times m}$-valued spectral density functions on the unit circle. Hence, $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ if and only if $\Phi^{-1} \in \mathcal{S}_+^{m\times m}(\mathbb{T})$. Given a Hermitian and positive-definite $n\times n$ matrix Σ, consider the problem of finding $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ that satisfies (3), i.e., that is compatible with Σ. This is a particular case of a *moment problem.* In the last ten years, much research has been produced, mainly by the Byrnes–Georgiou–Lindquist school, on generalized moment problems [1–5], and analytic interpolation with complexity constraint [6], and their applications to spectral estimation [7–9] and robust control [10]. It is worth recalling that two fundamental problems of control theory, namely the *covariance extension problem* and the *Nevanlinna–Pick interpolation problem of robust control*, can be recast in this form [5].

Eq. (3), where the unknown is Φ, is also a typical example of an *inverse* problem. Recall that a problem is said to be *well posed*, in the sense of Hadamard, if it admits a solution, such a solution is unique, and the solution depends continuously on the data. Inverse problems are typically *not* well posed. In our case, there may well be no solution Φ, and when a solution exists, there may be (infinitely) many. It was shown in [11], that the set of solutions is nonempty if and only if there exists $H \in \mathbb{C}^{m\times n}$ such that

$$\Sigma - A\Sigma A^* = BH + H^*B^*. \tag{4}$$

When (4) is feasible with $\Sigma > 0$, and there are infinitely many solutions Φ to (3). To select a particular solution it is natural to introduce an optimality criterion. For control applications, however, it is desirable that such a solution be of limited

☆ Partially supported by the Ministry of Education, University, and Research of Italy (MIUR), under Project 2006094843: *New techniques and applications of identification and adaptive control.*

∗ Corresponding author.
E-mail addresses: ramponif@control.ee.ethz.ch (F. Ramponi), augusto@dei.unipd.it (A. Ferrante), pavon@math.unipd.it (M. Pavon).

0167-6911/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sysconle.2010.01.001

complexity. It should namely be rational and with an *a priori* bound on its MacMillan degree. One of the great accomplishments of the Byrnes–Georgiou–Lindquist approach is having shown that the minimization of certain *entropy-like functionals* leads to solutions that satisfy this requirement. In [11], Georgiou provided an explicit expression for the spectrum $\hat{\Phi}$ that exhibits a *maximum entropy rate* among the solutions of (3).

Suppose now that some *a priori* information about Φ is available in the form of a spectrum $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$. Given G, Σ, and Ψ, we now seek a spectrum Φ, which is the closest to Ψ in a certain metric, among the solutions of (3). Paper [5] deals with such an optimization problem in the case when y is a scalar process. The criterion there is the Kullback–Leibler pseudo-distance from Ψ to Φ. A drawback of this approach is that it does not seem to generalize to the multivariable case. This motivated us to provide a suitable extension of the so-called *Hellinger distance* with respect to which the *multivariable* version of the problem is solvable (see [12,9]).

The main result of this paper is contained in Section 3. We show there that, under the feasibility assumption, the solution to the spectrum approximation problem with respect to both the scalar Kullback–Leibler pseudo-distance and the multivariable Hellinger distance depends continuously on Σ, thereby proving that these problems are well-posed. In Section 4 we deal with the case when only an *estimate* $\hat{\Sigma}$ of Σ is available. By applying the continuity results of Section 3, we prove a consistency result for the solutions to both approximation problems.

2. Spectrum approximation problems

In this section, we collect some background material on the spectrum approximation problems. The reader is referred to [11,5,12,9] for a more detailed treatment.

2.1. Feasibility of the moment problem

Let $\mathbb{H}(n)$ be the space of Hermitian $n \times n$ matrices, and $\mathcal{C}(\mathbb{T};\mathbb{H}(m))$ the space of $\mathbb{H}(m)$-valued continuous functions defined on the unit circle. Let the operator $\Gamma : \mathcal{C}(\mathbb{T};\mathbb{H}(m)) \to \mathbb{H}(n)$ be defined as follows:

$$\Gamma(\Phi) := \int G\Phi G^*. \tag{5}$$

Consider now the *range* of the operator Γ (as a vector space over the reals). We have the following result (see [9]).

Proposition 2.1. 1. *Let $\Sigma = \Sigma^* > 0$. The following are equivalent:*

- *There exists $H \in \mathbb{C}^{m\times n}$ which solves (4).*
- *There exists $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ such that $\int G\Phi G^* = \Sigma$.*
- *There exists $\Phi \in \mathcal{C}(\mathbb{T};\mathbb{H}(m))$, $\Phi > 0$ such that $\Gamma(\Phi) = \Sigma$.*

2. *Let $\Sigma = \Sigma^*$ (not necessarily definite). There exists $H \in \mathbb{C}^{m\times n}$ that solves (4) if and only if $\Sigma \in \text{Range}\ \Gamma$.*
3. *$X \in \text{Range}\ \Gamma^\perp$ if and only if $G^*(e^{j\vartheta})XG(e^{j\vartheta}) = 0\ \forall \vartheta \in [0, 2\pi]$.*

We define

$$P_\Gamma := \{\Sigma \in \text{Range}\ \Gamma \mid \Sigma > 0\}. \tag{6}$$

In view of Proposition 2.1, for each $\Sigma \in P_\Gamma$ the problem (3) is feasible.

2.2. Scalar approximation in the Kullback–Leibler pseudo-distance

In [5], the Kullback–Leibler pseudo-distance for spectral densities in $\mathcal{S}_+^{1\times 1}(\mathbb{T})$ was introduced:

$$\mathbb{D}(\Psi \parallel \Phi) = \int \Psi \log \frac{\Psi}{\Phi}. \tag{7}$$

As is well known, the corresponding quantity for probability densities originates in hypothesis testing, where it represents the mean information per observation for the discrimination of an underlying probability density from another [13]. The approximation problem goes as follows:

Problem 2.2. Given $\Sigma \in P_\Gamma$ and $\Psi \in \mathcal{S}_+^{1\times 1}(\mathbb{T})$, find Φ_o^{KL} that solves

$$\text{minimize} \quad \mathbb{D}(\Psi \parallel \Phi)$$
$$\text{over} \quad \left\{\Phi \in \mathcal{S}_+^{1\times 1}(\mathbb{T}) \mid \int G\Phi G^* = \Sigma\right\}. \tag{8}$$

Note that, following [5], and differently from optimization problems that are usual in the probability setting, we minimize (7) with respect to the *second* argument. The remarkable advantage of this approach is that, differently from an optimization with respect to the first argument, it will yield a *rational* solution whenever Ψ is rational. Let

$\mathcal{L}^{KL} := \{\Lambda \in \mathbb{H}(n) \mid G^*\Lambda G > 0, \quad \forall e^{i\vartheta} \in \mathbb{T}\}$.

For a given $\Lambda \in \mathcal{L}^{KL}$, consider the *Lagrangian functional*

$$L(\Phi;\Lambda) = \mathbb{D}(\Psi \parallel \Phi) + \left\langle \Lambda, \int G\Phi G^* - \Sigma \right\rangle, \tag{9}$$

where $\langle A, B\rangle := \text{tr}\,AB$ denotes the scalar product between the Hermitian matrices A and B. Observe that the term $\int G\Phi G^*$ between the brackets belongs to Range Γ by definition, while Σ belongs to Range Γ by the feasibility assumption. Hence, it is natural to restrict Λ to Range Γ, or, which is the same, to

$\mathcal{L}_\Gamma^{KL} := \mathcal{L}^{KL} \cap \text{Range}\ \Gamma$.

The functional (9) is *strictly convex* on $\mathcal{S}_+^{1\times 1}(\mathbb{T})$. Hence, its *unconstrained* minimization with respect to Φ can be pursued imposing that its derivative in an arbitrary direction $\delta\Phi$ is zero. This yields the form for the optimal spectrum:

$$\hat{\Phi}^{KL}(\Lambda) = \frac{\Psi}{G^*\Lambda G}. \tag{10}$$

As noted previously, inasmuch as Ψ is rational $\hat{\Phi}^{KL}(\Lambda)$ is also rational, and with a MacMillan degree less than or equal to $2n + \deg\Psi$. Now if $\hat{\Lambda} \in \mathcal{L}_\Gamma^{KL}$ is such that

$$\int G \frac{\Psi}{G^*\hat{\Lambda}G} G^* = \Sigma, \tag{11}$$

that is, if $\hat{\Lambda}$ is such that the corresponding optimal spectrum $\Phi_o^{KL} = \hat{\Phi}^{KL}(\hat{\Lambda})$ satisfies the constraint, then Φ_o^{KL} is the unique solution to the constrained approximation Problem 2.2. Finding such $\hat{\Lambda}$ is the objective of the *dual problem*, which is readily seen [5] to be equivalent to

$$\text{minimize} \quad \{J_\Psi^{KL}(\Lambda) \mid \Lambda \in \mathcal{L}_\Gamma^{KL}\} \tag{12}$$

where

$$J_\Psi^{KL}(\Lambda) = -\int \Psi \log G^*\Lambda G + \text{tr}\Lambda\Sigma. \tag{13}$$

This is also a convex optimization problem. The *existence* of a minimum is a highly nontrivial issue. Such an existence was proved in [5] resorting to a profound topological result, and in [14] by a less abstract argument.

Theorem 2.3. *The strictly convex functional J_Ψ^{KL} has a unique minimum point in $\mathcal{L}_\Gamma^{KL}$.*

The minimum point of the Theorem 2.3 provides the optimal solution to the primal Problem 2.2 via (10). Differently from the primal problem, whose domain $\mathcal{S}_+^{1\times 1}(\mathbb{T})$ is infinite-dimensional, the dual problem is finite-dimensional, hence the minimization of J_Ψ^{KL} can be accomplished with iterative numerical methods. The numerical minimization of J_Ψ^{KL} is not, however, a simple problem, because both the functional and its gradient are unbounded on $\mathcal{L}_\Gamma^{KL}$ (which is unbounded itself). Moreover, the reparameterization of $\mathcal{L}_\Gamma^{KL}$ may lead to a loss of convexity (see [5] and references therein). An alternative approach to this problem was proposed in [15,16].

2.3. Multivariable approximation in the Hellinger distance

In [12] the *Hellinger distance* between two spectral densities $\Phi, \Psi \in \mathcal{S}_+^{1\times 1}(\mathbb{T})$ was introduced:

$$d_H(\Phi,\Psi) := \left[\int \left(\sqrt{\Phi}-\sqrt{\Psi}\right)^2\right]^{1/2}. \quad (14)$$

As it happens for the Kullback–Leibler case, its counterpart for probability densities is well-known in mathematical statistics. Differently from the Kullback–Leibler case, this is a *bona fide* distance (note that (14) is nothing more that the L^2 distance between the square roots of Φ and Ψ, and that the square roots are particular instances of the *spectral factors*). A variational analysis similar to the one we have just seen is possible and leads to similar results. Let us focus directly on the multivariable extension of (14) that was developed in [12]. Given $\Phi, \Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$, we define the following quantity:

$$d_H(\Phi,\Psi) := \inf\left\{\|W_\Psi - W_\Phi\|_2 :\ W_\Psi, W_\Phi \in L_2^{m\times m},\ W_\Psi W_\Psi^* = \Psi,\ W_\Phi W_\Phi^* = \Phi\right\}. \quad (15)$$

Observe that $d_H(\Phi,\Psi)$ is simply the L^2 distance between the sets of *all the square spectral factors* of Φ and Ψ respectively. We have the following result (see [12]).

Theorem 2.4. *The following facts hold true:*

1. d_H *is a* bona fide *distance function.*
2. $d_H(\Phi,\Psi)$ *coincides with* (14) *when* Φ *and* Ψ *are scalar.*
3. *The infimum in* (15) *is indeed a minimum.*
4. *For any square spectral factor* $\bar{W}_\Psi$ *of* Ψ, *we have:*

$$d_H(\Phi,\Psi) = \inf_{W_\Phi}\left\{\|\bar{W}_\Psi - W_\Phi\|_2 :\ W_\Phi \in L_2^{m\times m}, W_\Phi W_\Phi^* = \Phi\right\}.$$

Fact 4 says that, if we fix a spectral factor of one spectrum and minimize only among the spectral factors of the other, the result is the same. Given $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ (and $G(z)n\times m$), we pose a minimization problem similar to Problem 2.2:

Problem 2.5. Given $\Sigma \in P_\Gamma$ and $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$, find Φ_o^H that solves

$$\text{minimize} \quad d_H(\Phi,\Psi)$$
$$\text{over} \quad \left\{\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T}) \mid \int G\Phi G^* = \Sigma\right\}. \quad (16)$$

In view of the facts 3 and 4 in the Theorem 2.4, once a spectral factor of Ψ is fixed, the same Problem 2.5 can be reformulated in terms of a minimization with respect to *spectral factors of* Φ:

Given $\Sigma \in P_\Gamma$ *and a spectral factor* W_Ψ *of* $\Psi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$, *find* W_o^H *that solves*

$$\text{minimize} \quad \operatorname{tr}\int (W_\Phi - W_\Psi)(W_\Phi - W_\Psi)^*$$
$$\text{over} \quad \left\{W_\Phi \in L_2^{m\times m} \middle| \int G W_\Phi W_\Phi^* G^* = \Sigma\right\}. \quad (17)$$

Consider the Lagrangian functional

$$H(W_\Phi,\Lambda) = \operatorname{tr}\int (W_\Phi - W_\Psi)(W_\Phi - W_\Psi)^* + \left\langle \Lambda, \int G W_\Phi W_\Phi^* G^* - \Sigma\right\rangle. \quad (18)$$

For the same reason as before, we restrict the matrix Λ to Range Γ. The functional (18) is *strictly convex* and, for a given Λ, its *unconstrained* minimization with respect to W_Φ yields the following condition for the optimal spectral factor $\hat{W}^H(\Lambda)$ (see [12] for details):

$$\hat{W}^H(\Lambda) - W_\Psi + G^*\Lambda G\,\hat{W}^H(\Lambda) = 0. \quad (19)$$

In order to ensure that the corresponding spectrum is integrable over the unit circle, we now require *a posteriori* that Λ belongs to the set

$$\mathcal{L}^H = \left\{\Lambda \in \mathbb{H}(n) \mid I + G^*\Lambda G > 0\ \forall e^{j\vartheta} \in \mathbb{T}\right\}$$

or, which is the same, that it belongs to the set

$$\mathcal{L}_\Gamma^H := \mathcal{L}^H \cap \text{Range}\ \Gamma. \quad (20)$$

Such a restriction yields the following optimal spectral factor and spectrum:

$$\begin{aligned}\hat{W}^H(\Lambda) &= (I+G^*\Lambda G)^{-1}W_\Psi,\\ \hat{\Phi}^H(\Lambda) &= \hat{W}^H(\Lambda)\hat{W}^{H^*}(\Lambda) = (I+G^*\Lambda G)^{-1}\Psi(I+G^*\Lambda G)^{-1}.\end{aligned} \quad (21)$$

Now if $\hat{\Lambda}$ is such that

$$\int G\,(I+G^*\hat{\Lambda}G)^{-1}\Psi(I+G^*\hat{\Lambda}G)^{-1}\,G^* = \Sigma, \quad (22)$$

then $W_o^H = \hat{W}^H(\hat{\Lambda})$ and $\Phi_o^H = \hat{\Phi}^H(\hat{\Lambda})$ are the unique solutions to the constrained approximation problems (17) and (16) respectively. In order to find such $\hat{\Lambda}$, one must solve the *dual problem*, which can be shown to be equivalent to

$$\text{minimize} \quad \{J_\Psi^H(\Lambda) \mid \Lambda \in \mathcal{L}_\Gamma^H\} \quad (23)$$

where

$$J_\Psi^H(\Lambda) = \operatorname{tr}\int (I+G^*\Lambda G)^{-1}\Psi + \operatorname{tr}\Lambda\Sigma. \quad (24)$$

The *existence* of a minimum is again a highly nontrivial issue. We have the following result (see [12]).

Theorem 2.6. *The strictly convex functional* J_Ψ^H *has a unique minimum point in* $\mathcal{L}_\Gamma^H$.

The minimum point of Theorem 2.6 provides the optimal solution to the primal Problem 2.5 via (21). It can be found by means of iterative numerical algorithms. The numerical minimization of J_Ψ^H is challenging due to reasons similar to the ones concerning J_Ψ^{KL}. In [9], we propose a matricial version of the Newton algorithm that avoids any reparameterization of $\mathcal{L}_\Gamma^H$, and proved its global convergence.

3. Well-posedness of the approximation problems

In this section, we show that both the dual problems (12) and (23) are well-posed, since their unique solution is continuous with respect to a small perturbation of Σ. The well-posedness of the respective *primal* problem then easily follows. All these continuity properties rely on the following basic result.

Theorem 3.1. *Let A be an open and convex subset of a finite-dimensional Euclidean space V. Let $f : A \to \mathbb{R}$ be a strictly convex function, and suppose that a minimum point $\bar{x}$ of f exists. Then, for all $\varepsilon > 0$, there exists $\delta > 0$ such that, for each $p \in \mathbb{R}^n$, $\|p\| < \delta$, the function $f_p : A \to \mathbb{R}$ defined as*

$$f_p(x) := f(x) - \langle p, x \rangle$$

admits an unique minimum point $\bar{x}_p$, and moreover

$$\|\bar{x}_p - \bar{x}\| < \varepsilon.$$

(Note: $f^*(p) := -f_p(\bar{x}_p)$ is the Fenchel dual *of f at p.)*

Proof. First, note that the minimum point $\bar{x}$ is unique, since f is strictly convex. Let $\varepsilon > 0$, and let $S(\bar{x}, \varepsilon) = \{\bar{x} + y \mid \|y\| = \varepsilon\}$ denote the sphere of radius ε centered in $\bar{x}$. Let moreover $B(\bar{x}, \varepsilon) = \{\bar{x} + y \mid \|y\| < \varepsilon\}$ denote the open ball of radius ε centered in $\bar{x}$ and $\bar{B}(\bar{x}, \varepsilon) = \{\bar{x} + y \mid \|y\| \le \varepsilon\}$ its closure. Then $\bar{B}(\bar{x}, \varepsilon) = B(\bar{x}, \varepsilon) \cup S(\bar{x}, \varepsilon)$, $\bar{B}(\bar{x}, \varepsilon)$ and $S(\bar{x}, \varepsilon)$ are compact, and $S(\bar{x}, \varepsilon)$ is the boundary of $B(\bar{x}, \varepsilon)$. Since f is continuous, it admits a minimum point $\bar{x} + y_\varepsilon$ over $S(\bar{x}, \varepsilon)$. Since $\bar{x}$ is the unique global minimum point of f, we must have $m_\varepsilon := f(\bar{x} + y_\varepsilon) - f(\bar{x}) > 0$. Then, for $\|y\| = \varepsilon$ we have

$$f(\bar{x} + y) - f(\bar{x}) \ge m_\varepsilon. \tag{25}$$

Let now $0 < \delta < m_\varepsilon/\varepsilon$. For $\|p\| < \delta$ and $\|y\| = \varepsilon$ we have

$$\langle p, y \rangle \le \|p\| \, \|y\| < \delta\varepsilon < m_\varepsilon \tag{26}$$

where the first inequality stems from the Cauchy–Schwartz inequality. From (25) and (26), we get for $\|y\| = \varepsilon$

$$f(\bar{x} + y) - f(\bar{x}) > \langle p, y \rangle = \langle p, \bar{x} + y \rangle - \langle p, \bar{x} \rangle$$
$$f_p(\bar{x} + y) > f_p(\bar{x})$$

that is,

$$f_p(x) > f_p(\bar{x})$$

for each $x \in S(\bar{x}, \varepsilon)$.

Now, since f is strictly convex and hence continuous, f_p is also strictly convex and continuous, and admits a minimum point $\bar{x}_p$ over the compact set $\bar{B}(\bar{x}, \varepsilon)$. But it follows from the previous considerations that such a minimum cannot belong to $S(\bar{x}, \varepsilon)$. Hence, it must belong to the open ball $B(\bar{x}, \varepsilon)$. *As such, $\bar{x}_p$ is also a local minimum of f_p over A, but since f_p is strictly convex, it is also the unique global minimum point*. Summing up, for fixed $\varepsilon > 0$, there exists $\delta > 0$ such that, if $\|p\| < \delta$, then f_p admits a unique minimum $\bar{x}_p$ over A. It follows from the previous analysis that, for sufficiently small δ, $\bar{x}_p$ belongs to $B(\bar{x}, \varepsilon)$. This proves the theorem. □

3.1. Well-posedness of Kullback–Leibler approximation

Consider the dual functional (13), and let us make its dependence upon Σ explicit:

$$J_\Psi^{KL}(\Lambda; \Sigma) = -\int \Psi \log G^* \Lambda G + \operatorname{tr} \Lambda \Sigma.$$

J_Ψ^{KL} is a strictly convex functional over $\mathcal{L}_\Gamma^{KL}$, which is an open and convex subset of the Euclidean space Range Γ. Due to Theorem 2.3, it does admit a minimum point

$$\hat{\Lambda}^{KL}(\Sigma) = \arg\min_\Lambda J_\Psi^{KL}(\Lambda; \Sigma).$$

In the latter, we have made the dependence of $\hat{\Lambda}^{KL}$ on Σ explicit. We proceed next to study this dependence. Let $\delta\Sigma$ be a perturbation of Σ. We have

$$J_\Psi^{KL}(\Lambda; \Sigma + \delta\Sigma) = -\int \Psi \log G^* \Lambda G + \operatorname{tr} \Lambda \Sigma + \operatorname{tr} \Lambda \delta\Sigma$$
$$= J_\Psi^{KL}(\Lambda; \Sigma) + \langle \delta\Sigma, \Lambda \rangle .$$

It follows from Theorem 3.1, where the role of $\delta\Sigma$ is played by $-p$, that for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $\|\delta\Sigma\|_F < \delta$, then $J_\Psi^{KL}(\Lambda; \Sigma + \delta\Sigma)$ again admits a minimum point

$$\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) = \arg\min_\Lambda J_\Psi^{KL}(\Lambda; \Sigma + \delta\Sigma) \tag{27}$$

and the distance $\|\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) - \hat{\Lambda}^{KL}(\Sigma)\|_F$ is less than ε. The above observation implies the well-posedness of the dual problem:

Corollary 3.2. *The map $\hat{\Lambda}^{KL} : P_\Gamma \to \mathcal{L}_\Gamma^{KL}$ is continuous.*

Consider now the primal problem. The variational analysis yielded the following optimal solution, where the dependence upon Σ has been made explicit:

$$\Phi_o^{KL}(\Sigma) = \hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma)) = \frac{\Psi}{G^* \, \hat{\Lambda}^{KL}(\Sigma) \, G}.$$

We have the following result.

Theorem 3.3. *The map $\Phi_o^{KL} : P_\Gamma \to L_\infty$ is continuous.*

Proof. Recall that $\hat{\Lambda}^{KL}(\Sigma)$ is the solution of the dual problem where the true asymptotic state variance is known, and let $\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma)$ be the solution to the dual problem with respect to a perturbed covariance. Let $\hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma))$ and $\hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma))$ be the corresponding solutions to the primal problem. Then

$$\|\hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma)) - \hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma))\|_\infty$$
$$= \left\| \frac{\Psi}{G^* \, \hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) \, G} - \frac{\Psi}{G^* \, \hat{\Lambda}^{KL}(\Sigma) \, G} \right\|_\infty$$
$$\le \|\Psi\|_\infty \left\| \frac{1}{G^* \, \hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) \, G} - \frac{1}{G^* \, \hat{\Lambda}^{KL}(\Sigma) \, G} \right\|_\infty .$$

It is easily seen that for each $\eta > 0$ we can choose $\varepsilon > 0$ such that if $\|\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) - \hat{\Lambda}^{KL}(\Sigma)\|_F < \varepsilon$, then

$$\max_\vartheta |G^* \hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) G - G^* \hat{\Lambda}^{KL}(\Sigma) G|$$
$$= \max_\vartheta |G^\top(e^{-j\vartheta})(\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma) - \hat{\Lambda}^{KL}(\Sigma)) G(e^{j\vartheta})| < \eta.$$

Finally, from the above observation, from the Corollary 3.2, and from the continuity of the function $\frac{1}{x}$ over $\mathbb{R}^+$, it follows that for each $\mu > 0$, there exists $\delta > 0$ such that, for all $\|\delta\Sigma\|_F < \delta$, $\|\hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma + \delta\Sigma)) - \hat{\Phi}^{KL}(\hat{\Lambda}^{KL}(\Sigma))\|_\infty < \mu$. □

Corollary 3.4. *The problem*

$$\arg\min_\Phi D(\Psi \parallel \Phi) \quad \text{such that} \quad \int G \Phi G^* = \Sigma$$

is well-posed for $\Sigma \in P_\Gamma$ and for variations $\delta\Sigma$ that belong to Range *Γ.*

3.2. Well-posedness of the Hellinger approximation

Consider the dual functional (24):

$$J_\Psi^H(\Lambda; \Sigma) = \operatorname{tr} \int (I + G^* \Lambda G)^{-1} \Psi + \operatorname{tr} \Lambda \Sigma.$$

J_Ψ^H is a strictly convex functional over $\mathcal{L}_\Gamma^H$, which is an open and convex subset of the Euclidean space Range Γ. Due to the Theorem 2.3, it admits a minimum point

$$\hat{\Lambda}^H(\Sigma) = \arg\min_\Lambda J_\Psi^H(\Lambda; \Sigma).$$

Let as before $\delta\Sigma$ be a perturbation of Σ. Then

$$J_\Psi^H(\Lambda; \Sigma + \delta\Sigma) = J_\Psi^H(\Lambda; \Sigma) + \langle \delta\Sigma, \Lambda \rangle .$$

Theorem 3.1 implies the following

Corollary 3.5. *The map* $\hat{\Lambda}^H : P_\Gamma \to \mathcal{L}^H_\Gamma$ *is continuous.*

The variational analysis yielded the optimal solution for the primal problem

$$\Phi^H_o(\Sigma) = \hat{\Phi}^H(\hat{\Lambda}^H(\Sigma)) = (I + G^* \hat{\Lambda}^H(\Sigma)\, G)^{-1} \times \Psi (I + G^* \hat{\Lambda}^H(\Sigma)\, G)^{-1}, \quad (28)$$

and considerations similar to those of the Theorem 3.3 lead to the following

Theorem 3.6. *The map* $\Phi^H_o : P_\Gamma \to L^{m\times m}_\infty$ *is continuous.*

To prove Theorem 3.6 we exploit the following result established in [9] (Lemma 5.2):

Lemma 3.7. *Define* $Q_\Lambda(z) = I + G^*(z)\Lambda G(z)$. *Consider a sequence* $\Lambda_n \in \mathcal{L}^H_\Gamma$ *converging to* $\Lambda \in \mathcal{L}^H_\Gamma$. *Then* $Q^{-1}_{\Lambda_n}$ *are well defined and continuous on* $\mathbb{T}$ *and converge* uniformly *to* Q^{-1}_Λ *on* $\mathbb{T}$.

Proof of Theorem 3.6. Let $Q_\Lambda(z; \Sigma) = I + G^*(z)\, \hat{\Lambda}^H(\Sigma)\, G(z)$. Apply the Corollary 3.5 and Lemma 3.7 to establish the continuity of the map from P_Γ to $L^{m\times m}_\infty$ defined by $\Sigma \mapsto Q^{-1}_\Lambda$. The continuity of Φ^H_o follows from the continuity of the matrix multiplication. □

Corollary 3.8. *The problem*

$$\arg\min_\Phi d_H(\Phi, \Psi) \quad \text{such that} \quad \int G\Phi G^* = \Sigma$$

is well-posed, for $\Sigma \in P_\Gamma$ *and for variations* $\delta\Sigma$ *that belong to* Range Γ.

4. Consistency

So far we have shown that both the approximation problems admit a unique solution for all $\Sigma \in P_\Gamma$, and that the solution is continuous with respect to the variations $\delta\Sigma \in$ Range Γ. The necessity of a restriction to Range Γ becomes crucial in the case when we only have an *estimate* $\hat{\Sigma}$ of Σ.

In line with the Byrnes–Georgiou–Lindquist theory, and following an estimation procedure we have sketched in [9], we want to use the above theory to provide an estimate $\hat{\Phi}$ of the true spectrum of the process y.

Let $G(z)$ and Ψ be given. Suppose that we feed $G(z)$ with a *finite* sequence of observations, say $\{y_1, \ldots, y_N\}$ of the process. Observing the states of the system, say $\{x_1, \ldots, x_N\}$, we then compute a Hermitian and positive definite *estimate* $\hat{\Sigma}$ of the asymptotic state covariance, such as

$$\hat{\Sigma} = \frac{1}{N}\sum_{k=1}^{N} x_k x_k^*.$$

This is provably consistent, and also unbiased, for we have supposed from the beginning that y has a zero mean. We seek an estimate $\hat{\Phi}$ of Φ by solving an approximation problem with respect to $G(z)$, Ψ, and $\hat{\Sigma}$.

Since $\hat{\Sigma}$ is not the true variance any longer, the constraint (3) may be not feasible. Hence, in order to find a solution $\hat{\Phi}$, we need to find a second estimate $\bar{\Sigma}$, close to the first, such that the (4) is feasible with the covariance matrix $\bar{\Sigma}$. A reasonable way to proceed is to let $\bar{\Sigma}$ be the projection of $\hat{\Sigma}$ onto Range Γ. Since the orthogonal projectors from $\mathbb{H}(n)$ to a subspace of the $\mathbb{H}(n)$ are continuous functions, if $\hat{\Sigma}(x_1, \ldots, x_N)$ is a consistent estimator of Σ, then $\bar{\Sigma}$ is also a consistent estimator of Σ.

The problem that *may* come up in proceeding this way is that the projection onto Range Γ needs not be positive definite (that is, it may not belong to P_Γ), even if $\hat{\Sigma}$ is. If this is the case, the correct procedure to estimate Σ while preserving the structure of a state covariance compatible with $G(z)$ is to find $\bar{\Sigma} \in P_\Gamma$ which is the closest to $\hat{\Sigma}$ in a suitable distance. This is an optimization problem in itself.

The continuity results of the preceding sections imply two strong consistency results. Let $\bar{\Sigma}(x_1, \ldots, x_N) \in P_\Gamma$ denote a consistent estimator of Σ. Let $\Phi^{KL}_o(\Sigma)$ be the solution to the Kullback–Leibler approximation problem with respect to the true asymptotic variance and $\Phi^{KL}_o(\bar{\Sigma}(x_1, \ldots, x_N))$ be the solution of the same problem with respect to the estimate.

Corollary 4.1. *If*

$$\lim_{N\to\infty} \bar{\Sigma}(x_1, \ldots, x_N) = \Sigma \quad \text{a.s.}, \quad (29)$$

then

$$\lim_{N\to\infty} \|\Phi^{KL}_o(\bar{\Sigma}(x_1, \ldots, x_N)) - \Phi^{KL}_o(\Sigma)\|_\infty = 0 \quad \text{a.s.}$$

Proof. From the continuity of the map Φ^{KL}_o we have that, excepting a set of zero probability,

$$\lim_{N\to\infty} \Phi^{KL}_o\left(\bar{\Sigma}(x_1(\omega), \ldots, x_N(\omega))\right) = \Phi^{KL}_o\left(\lim_{N\to\infty} \bar{\Sigma}(x_1(\omega), \ldots, x_N(\omega))\right) = \Phi^{KL}_o(\Sigma),$$

where the first limit is taken in $L_\infty(\mathbb{T})$. □

As for the Hellinger multivariable approximation problem, let $\Phi^H_o(\Sigma)$ be the solution with respect to the true asymptotic variance and $\Phi^H_o(\bar{\Sigma}(x_1, \ldots, x_N))$ be the solution with respect to the estimate. Employing the very same technique used for the proof of Corollary 4.1 it is easy to establish the following consistency result for the problem associated to the multivariable Hellinger distance.

Corollary 4.2. *If*

$$\lim_{N\to\infty} \bar{\Sigma}(x_1, \ldots, x_N) = \Sigma \quad \text{a.s.},$$

then

$$\lim_{N\to\infty} \|\Phi^H_o(\bar{\Sigma}(x_1, \ldots, x_N)) - \Phi^H_o(\Sigma)\|_\infty = 0 \quad \text{a.s.}$$

5. Conclusion and future work

In this paper, we have considered the constrained spectrum approximation problems with respect to both the Kullback–Leibler pseudo-distance (scalar case) and the Hellinger distance (multivariable case). The range of the operator $\Gamma : \Phi \mapsto \int G\Phi G^*$ is the subspace of the Hermitian matrices that conveys all the structure that is needed from a positive-definite matrix in order to be an asymptotic covariance matrix of the system with transfer function $G(z)$. As such, it is also a natural subspace to which the domains of the respective dual problems should be constrained. We have shown that the condition $\Sigma \in P_\Gamma$ is not only necessary for the feasibility of the moment problem $\{\Phi \mid \int G\Phi G^* = \Sigma\}$, but is also sufficient for the continuity of the respective solutions with respect to Σ. This fact implies well-posedness of both kinds of approximation problem, and implies the consistency of the respective solutions with respect to a consistent estimator $\hat{\Sigma}$ of Σ, as long as it is restricted to Range Γ. Similar results can be established along the same lines when employing *any* other (pseudo-)distance, as long as the functional form of the primal optimum depends continuously upon the Lagrange parameter Λ.

As suggested by an anonymous reviewer, to whom we are grateful, it would be interesting to investigate the possibility of establishing the (local) moduli of continuity of the maps Φ_o^{KL} and Φ_o^{H}. In particular, it would be worthwhile to study this issue when Σ tends to become singular and/or the prior spectrum Ψ may have roots on the unit circle. The complete analysis appears to be challenging and we defer this investigation to future research. Here, we shall be content with making an observation on the *maximum entropy solution*, i.e., the Kullback–Leibler optimal solution in the case when $\Psi = I$. It has been shown in [2] that in this important case the map $\hat{\Lambda}^{KL}(\Sigma)$ may be explicitly written as

$$\hat{\Lambda}^{KL}(\Sigma) = \Sigma^{-1}B\left(B^*\Sigma^{-1}B\right)^{-1}B^*\Sigma^{-1}.$$

It is apparent that this function is locally Lipschitz-continuous but the Lipschitz constant may diverge as Σ tends to become singular.

References

[1] C.I. Byrnes, S. Gusev, A. Lindquist, A convex optimization approach to the rational covariance extension problem, SIAM J. Control and Optim. 37 (1999) 211–229.
[2] T. Georgiou, Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parameterization, IEEE Trans. Automat. Control 47 (2002) 1811–1823.
[3] C.I. Byrnes, A. Lindquist, The generalized moment problem with complexity constraint, Integral Equations Operator Theory 56(2) (2006) 163–180.
[4] T. Georgiou, Relative entropy and the multivariable multidimensional moment problem, IEEE Trans. Inform. Theory 52 (2006) 1052–1066.
[5] T. Georgiou, A. Lindquist, Kullback–Leibler approximation of spectral density functions, IEEE Trans. Inform. Theory 49 (2003) 2910–2917.
[6] A. Blomqvist, A. Lindquist, R. Nagamune, Matrix-valued Nevanlinna–Pick interpolation with complexity constraint: An optimization approach, IEEE Trans. Automat. Control 48 (2003) 2172–2190.
[7] C.I. Byrnes, T. Georgiou, A. Lindquist, A new approach to spectral estimation: A tunable high-resolution spectral estimator, IEEE Trans. Signal Process. 49 (2000) 3189–3205.
[8] T. Georgiou, A. Lindquist, A convex optimization approach to ARMA modeling, IEEE Trans. Automat. Control AC-53 (2008) 1108–1119.
[9] F. Ramponi, A. Ferrante, M. Pavon, A globally convergent matricial algorithm for multivariate spectral estimation, IEEE Trans. Automat. Control 54 (10) (2009) 2376–2388.
[10] T. Georgiou, A. Lindquist, Remarks on control design with degree constraint, IEEE Trans. Automat. Control AC-51 (2006) 1150–1156.
[11] T. Georgiou, The structure of state covariances and its relation to the power spectrum of the input, IEEE Trans. Automat. Control 47 (2002) 1056–1066.
[12] A. Ferrante, M. Pavon, F. Ramponi, Hellinger vs. Kullback–Leibler multivariable spectrum approximation, IEEE Trans. Automat. Control 53 (2008) 954–967.
[13] S. Kullback, Information Theory and Statistics, 2nd ed., Dover, Mineola, NY, 1968.
[14] A. Ferrante, M. Pavon, F. Ramponi, Further results on the Byrnes-Georgiou-Lindquist generalized moment problem, in: A. Chiuso, A. Ferrante, S. Pinzoni (Eds.), Modeling, Estimation and Control: Festschrift in Honor of Giorgio Picci on the Occasion of his Sixty-Fifth Birthday, Springer-Verlag, 2007, pp. 73–83.
[15] M. Pavon, A. Ferrante, On the Georgiou-Lindquist approach to constrained Kullback–Leibler approximation of spectral densities, IEEE Trans. Automat. Control 51 (2006) 639–644.
[16] A. Ferrante, F. Ramponi, F. Ticozzi, On the convergence of an efficient algorithm for Kullback–Leibler approximation of spectral densities, IEEE Trans. Automat. Control (2009) (submitted for publication) arXiv:0911.0434v1 [math.OC].