Further Results on the Byrnes-Georgiou-Lindquist Generalized Moment Problem*

Augusto Ferrante[1], Michele Pavon[2], and Federico Ramponi[1]

[1] Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy
augusto@dei.unipd.it, rampo@dei.unipd.it
[2] Dipartimento di Matematica Pura ed Applicata, Università di Padova, via Trieste 63, 35131 Padova Italy
pavon@math.unipd.it

Summary. In this paper, we consider the problem of finding, among solutions of a moment problem, the best Kullback-Leibler approximation of a given *a priori* spectral density. We present a new complete existence proof for the dual optimization problem in the Byrnes-Lindquist spirit. We also prove a descent property for a matricial iterative method for the numerical solution of the dual problem. The latter has proven to perform extremely well in simulation testbeds.

1 Introduction

The concept of *positive real function*, originating in Networks Theory with Brune in 1930, has proven to be one of the deepest and most unifying ones of electrical engineering. Names such as Foster, Cauer, Bode, Darlington, Youla, Popov, Kalman, Yakubovich, Faurre, B.D.O. Anderson, J.C. Willems spring to mind. The quest first posed by Kalman in 1964 for a realization theory for stochastic systems could rely on these precious foundations. The strict sense version of the *stochastic realization problem* (also called *Markovian representation problem*) has roots also in the theory of Markov processes and in mathematical statistics (Bahadur transitive sufficient statistics, etc.). Giorgio Picci was one of the pioneers (together with McKean [39], Akaike [1, 2] and Ruckebusch [46]) in the early-mid seventies in this field. Perhaps, among the forerunners, he was the only one that drew equal inspiration from both of these areas, due to the fact that he had equal interest in the concepts of Systems Theory and of Mathematical Statistics [43, 44, 45]. Perhaps, this can be recognized as one of the main characteristics of all of Giorgio's rather diversified scientific production, which lays at the interface between stochastic systems theory and mathematical statistics. Another salient aspect of his research work is the taste for profound, foundational questions that continues today, for instance, in his work on subspace methods identification.

One of us (M.P.) had the privilege to witness the early days of the great Lindquist-Picci collaboration [33]- [38], and to receive continuous, generous help from Giorgio

* Work partially supported by the MIUR-PRIN Italian grant "New Techniques and Applications of Identification and Adaptive Control".

A. Chiuso et al. (Eds.): Modeling, Estimation and Control, LNCIS 364, pp. 73–83, 2007.
springerlink.com

both as a student and as a young researcher. The best way I can pay a personal tribute to Giorgio as a scientist is to observe that, although I have interacted and continue to interact with a large number and spectrum of colleagues, whenever there is a deep question on the table, I still go or address people to Giorgio, precisely as I did more than thirty years ago.

It is a joy to celebrate his devotion to science and research *per se* that continues unabated to this day, and his manifold, benchmark contributions to stochastic systems theory.

In this paper, we study a generalized moment problem for spectral densities in the spirit of Byrnes-Georgiou-Lindquist [4, 6, 7, 8, 9, 10, 11, 17, 21, 22, 23, 24, 25, 26, 27, 40]. This problem includes as special case the *covariance extension problem* and Nevanlinna-Pick interpolation for positive real functions. It also includes as special case a maximum entropy problem which has been shown to be related to a special one-step ahead Wiener-Kolmogorov prediction problem [27]. It features a Kullback-Leibler type index. It lays, therefore, very much in the center of the field in which Giorgio has been active for some forty years.

In the Byrnes-Georgiou-Lindquist approach, the smooth parametrization of the solutions to the generalized moment problem occurs in a convex optimization setting. The Kullback-Leibler criterion, where optimization is performed with respect to the second argument, is employed as cost index. Other distances between spectra have been recently investigated in [18, 19, 28, 29]. The contribution of this paper is twofold: On the one hand we provide a detailed and complete existence proof for the dual optimization problem (Section 5) in the Lindquist-Byrnes spirit [11]. On the other hand, we show that the matricial algorithm proposed in [41] for the numerical solution of the dual problem may be viewed as a modified steepest descent method (Section 6).

2 A Generalized Moment Problem

Let $\mathcal{S}_+(\mathbb{T})$ be the family of bounded, coercive, spectral density functions on the unit circle. Thus, a measurable, bounded function Φ belongs to $\mathcal{S}_+(\mathbb{T})$ if the values of Φ are real and bounded away from zero. Notice that $\Phi \in \mathcal{S}_+(\mathbb{T})$ if and only if $\Phi^{-1} \in \mathcal{S}_+(\mathbb{T})$. Let $\Psi \in \mathcal{S}_+(\mathbb{T})$ represent an *a priori* estimate of the spectrum of an underlying zero-mean, wide-sense stationary stochastic process $\{y(k), k \in \mathbb{Z}\}$. Suppose we can estimate the asymptotic covariance Σ of the n-dimensional stationary process $\{x_k; k \in \mathbb{Z}\}$ satisfying

$$x_{k+1} = Ax_k + By_k, \quad k \in \mathbb{Z}, \tag{1}$$

where A is a *stability matrix* and the pair (A, B) is reachable. The rational transfer function

$$G(z) = (zI - A)^{-1}B, \qquad A \in \mathbb{C}^{n\times n},\ B \in \mathbb{C}^{n\times 1}, \tag{2}$$

models a bank of filters. When Ψ is not consistent with Σ, we need to find $\Phi \in \mathcal{S}_+(\mathbb{T})$ that is closest to Ψ in a suitable sense among spectra satisfying

$$\int G(e^{i\vartheta})\Phi(e^{i\vartheta})G^*(e^{i\vartheta}) = \Sigma, \tag{3}$$

where star denotes transposition plus conjugation. The Hermitian matrix Σ is assumed to be positive definite and integration takes place on $[-\pi, \pi]$ with respect to the normalized Lebesgue measure $d\vartheta/2\pi$.

The question of existence of $\Phi \in \mathcal{S}_+(\mathbb{T})$ satisfying (3) and, when existence is granted, the parametrization of all solutions to (3), may be viewed as a generalized moment problem. We refer to [30] and references therein for a full discussion on the importance and on the manifold applications of such problem. Here, we only mention that, by suitably choosing the matrices A and B, this problem reduces the celebrated *covariance extension problem*, see [30] for details. Existence of $\Phi \in \mathcal{S}_+(\mathbb{T})$ satisfying constraint (3) is a nontrivial issue. It was shown in [23,24] that such family is nonempty if and only if there exists $H \in \mathbb{C}^{1\times n}$ such that

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

For simplicity of notation, we reformulate the constraint by normalizing Σ to the identity. Indeed, the set of solutions to (3) does not change if we replace (Σ, A, B) with $(I, A' = \Sigma^{-1/2} A \Sigma^{1/2}, B' = \Sigma^{-1/2} B)$. Notice that in this way G is replaced by $G' := \Sigma^{-1/2} G$. Thus constraint (3) from now on reads

$$\int G\Phi G^* = I. \tag{5}$$

3 Kullback-Leibler Criterion

In [30], the following problem is considered:

Problem 1. Given $\Psi \in \mathcal{S}_+(\mathbb{T})$, find $\hat{\Phi}$ that solves

$$\text{minimize} \quad \mathbb{S}(\Psi \| \Phi) \tag{6}$$

$$\text{over} \quad \left\{ \Phi \in \mathcal{S}(\mathbb{T}) \mid \int G\Phi G^* = I \right\}, \tag{7}$$

where $\mathbb{S}(\Psi\|\Phi)$ is the Kullback-Leibler index:

$$\mathbb{S}(\Psi\|\Phi) = \int \Psi \log\left(\frac{\Psi}{\Phi}\right).$$

As is well known, this pseudo-distance originates in hypothesis testing, where it represents the mean information for observation for discrimination of an underlying probability density from another [32, p.6]. It also plays a central role in information theory, identification, stochastic processes, etc., see e.g. [3,12,13,15,20,31,42,48] and references therein. It is also known in these fields as *divergence*, *relative entropy*, *information distance*. etc. If

$$\int \Phi = \int \Psi,$$

we have $\mathbb{S}(\Psi\|\Phi) \geq 0$. The choice of $\mathbb{S}(\Psi\|\Phi)$ as a distance measure, even for spectra that have different zeroth moment, is discussed in [30, Section III]. This choice is essentially based on the possibility of rescaling Ψ. Clearly, in this way the optimization problem amounts to approximating the "shape" of the a priori spectrum. In this spirit, Georgiou has recently investigated other distances between *rays* of power spectra, [28,29]. This is of course sensible to pursue in several engineering applications such as speech processing or prediction problems.

4 Optimality Conditions and the Dual Problem

The variational analysis in [30] is outlined as follows (see also [41]). Let

$$\mathcal{L}'_+ := \{\Lambda = \Lambda^* \in \mathbb{C}^{n\times n} : \ G^*\Lambda G > 0, \forall e^{i\vartheta} \in \mathbb{T}\}. \tag{8}$$

For $\Lambda \in \mathcal{L}'_+$ consider the *Lagrangian function*

$$\begin{aligned} L(\Phi,\Lambda) &= \mathbb{S}(\Psi\|\Phi) + \mathrm{tr}\left(\Lambda\left(\int G\Phi G^* - I\right)\right) \\ &= \mathbb{S}(\Psi\|\Phi) + \int G^*\Lambda G\Phi - \mathrm{tr}\,(\Lambda), \end{aligned} \tag{9}$$

where "tr" denotes the trace operator. Consider the *unconstrained* minimization of the strictly convex functional $L(\Phi,\Lambda)$:

$$\mathrm{minimize}\{L(\Phi,\Lambda) | \Phi \in \mathcal{S}(\mathbb{T})\} \tag{10}$$

This is a convex optimization problem.

Theorem 1. *Suppose $\hat{\Lambda} \in \mathcal{L}'_+$ is such that*

$$\int G\frac{\Psi}{G^*\hat{\Lambda}G}G^* = I. \tag{11}$$

Then $\hat{\Phi}$ given by

$$\hat{\Phi} = \frac{\Psi}{G^*\hat{\Lambda}G} \tag{12}$$

is the unique solution of Problem 1.

Thus, the original Problem 1 is now reduced to finding $\hat{\Lambda} \in \mathcal{L}'_+$ satisfying (11). We define the linear operator $\Xi : \mathcal{H} \rightarrow \mathcal{C}(\mathbb{T})$ by $\Xi(\Lambda) = G^*\Lambda G$, where $\mathcal{H}$ is the set of Hermitian matrices of dimension $n \times n$ and $\mathcal{C}(\mathbb{T})$ is the set of continuous functions on $\mathbb{T}$. Observe that if $\Lambda \in \mathcal{L}'_+$ satisfies (11), then, for all $\Lambda_K \in \ker \Xi$, the sum $\Lambda + \Lambda_K$ also belongs to $\mathcal{L}'_+$ and satisfies (11). Hence, we may restrict the search for $\hat{\Lambda}$ to the orthogonal complement of $\ker \Xi$ i.e. to the range of the adjoint operator $\Gamma := \Xi^*$. It is easy to see that the operator Γ is defined by

$$\Gamma(\Phi) = \int G\Phi G^*, \qquad \Phi \in \mathcal{C}(\mathbb{T}). \tag{13}$$

Thus, by setting

$$\mathcal{L}_+ := \mathcal{L}'_+ \cap [\ker(\Xi)]^\perp = \mathcal{L}'_+ \cap \mathrm{Range}(\Gamma) \tag{14}$$

we are reduced to find $\hat{\Lambda} \in \mathcal{L}_+$ satisfying (11). This is accomplished via duality theory. Consider the dual functional

$$\Lambda \to \inf\{L(\Phi, \Lambda) | \Phi \in \mathcal{S}(\mathbb{T})\}.$$

For $\Lambda \in \mathcal{L}_+$, the dual functional takes the form

$$\Lambda \to L\left(\frac{\Psi}{G^*\Lambda G}, \Lambda\right) = \int \Psi \log G^*\Lambda G - \mathrm{tr}\,(\Lambda) + \int \Psi. \tag{15}$$

Consider now the maximization of the dual functional (15) over the set $\mathcal{L}_+$. Let, as in [30],

$$J_\Psi(\Lambda) := -\int \Psi \log G^*\Lambda G + \mathrm{tr}\,(\Lambda). \tag{16}$$

The dual problem is then equivalent to

$$\text{minimize} \quad \{J_\Psi(\Lambda) | \Lambda \in \mathcal{L}_+\}. \tag{17}$$

The dual problem is a strictly convex optimization problem [30]. Byrnes and Lindquist have shown in [11] that J_Ψ has a unique minimum point in $\mathcal{L}_+$. This result implies that, under assumption (4), there exists a (unique) $\hat{\Lambda} \in \mathcal{L}_+$ satisfying (11). Such a $\hat{\Lambda}$ then provides the optimal solution of the primal problem (6)-(7) via (12). In the next section, we provide a new detailed proof of this result inspired by that of [11].

5 An Existence Theorem

Let us consider the closure of $\mathcal{L}_+$, given by

$$\overline{\mathcal{L}}_+ = \{\Lambda = \Lambda^* \in \mathbb{C}^{n\times n} : \Lambda \in \mathrm{Range}(\Gamma),\ G^*\Lambda G \geq 0, \forall e^{i\vartheta} \in \mathbb{T}\}. \tag{18}$$

On the convex set $\overline{\mathcal{L}}_+$, we define the sequence of functions

$$J_\Psi^n(\Lambda) := \mathrm{tr}\,(\Lambda) - \int \Psi \log\left(G^*\Lambda G + \frac{1}{n}\right). \tag{19}$$

Lemma 1. *The pointwise limit $J_\Psi^\infty(\Lambda) = \lim_n J_\Psi^n(\Lambda)$ exists and defines a lower semi-continuous, convex function on $\overline{\mathcal{L}}_+$ with values in the extended reals.*

Proof. For each n, J_Ψ^n is a continuous, convex function on the closed convex set $\overline{\mathcal{L}}_+$. Hence epi (J_Ψ^n), the epigraph of J_Ψ^n, is a closed, convex subset of $\mathbb{C}^{n\times n}\times\mathbb{R}$. Moreover, for $\Lambda \in \overline{\mathcal{L}}_+$, $J_\Psi^n(\Lambda) < J_\Psi^{n+1}(\Lambda)$. Hence, J_Ψ^∞ is well defined and in fact $J_\Psi^\infty(\Lambda) = \sup_n J_\Psi^n(\Lambda)$. It follows that epi $(J_\Psi^\infty) = \cap_n$epi (J_Ψ^n) is also closed and convex. We conclude that J_Ψ^∞ is lower semicontinuous and convex on $\overline{\mathcal{L}}_+$. ■

Lemma 2. *Assume (4). Then,*

1. J_Ψ^∞ *is bounded below on* $\overline{\mathcal{L}}_+$;
2. $J_\Psi^\infty(\Lambda) = J_\Psi(\Lambda)$ *on* $\mathcal{L}_+$;
3. $J_\Psi^\infty(\Lambda)$ *is finite on all of* $\overline{\mathcal{L}}_+ \setminus \{0\}$.

Proof. By (4), there exists $\Phi_1 \in \mathcal{S}_+(\mathbb{T})$ satisfying (5), namely $\int G\Phi_1 G^* = I$. Hence, $\mathrm{tr}\,(\Lambda)$ can be written as $\mathrm{tr}\,(\Lambda \int G\Phi_1 G^*) = \int G^*\Lambda G\Phi_1$, and we get

$$J_\Psi^n(\Lambda) = \int \left[G^*\Lambda G\Phi_1 - \Psi \log\left(G^*\Lambda G + \frac{1}{n}\right)\right]$$
$$= \int \Phi_1 \left[G^*\Lambda G - \frac{\Psi}{\Phi_1}\log\left(G^*\Lambda G + \frac{1}{n}\right)\right].$$

Since the function $x - \beta\log(x + \frac{1}{n})$ with $\beta > 0$ attains its minimum at $x = \beta - \frac{1}{n}$, we get

$$J_\Psi^n(\Lambda) = \int \Phi_1 \left[G^*\Lambda G - \frac{\Psi}{\Phi_1}\log\left(G^*\Lambda G + \frac{1}{n}\right)\right] \geq \int \psi - \frac{1}{n}\int \Phi_1 - \mathbb{S}(\Psi||\Phi_1).$$

We conclude that $J_\Psi^\infty \geq \int \psi - \mathbb{S}(\Psi||\Phi_1)$ on all of $\overline{\mathcal{L}}_+$. To establish 2, notice that, by Beppo Levi's theorem,

$$J_\Psi^\infty(\Lambda) := \mathrm{tr}\,(\Lambda) - \int \lim_{n\to\infty} \Psi \log\left(G^*\Lambda G + \frac{1}{n}\right), \quad \Lambda \in \overline{\mathcal{L}}_+. \tag{20}$$

To prove 3, observe that for $0 \neq \Lambda \in \partial\mathcal{L}_+$, the boundary of $\mathcal{L}_+$, $G^*\overline{\Lambda}G$ is a nonzero rational spectral density so that $\log G^*\overline{\Lambda}G$ is integrable over $\mathbb{T}$ [47, pag. 64]. Since Ψ is bounded, also $\Psi\log G^*\overline{\Lambda}G$ is integrable. ∎

In view of these lemmata, we extend $J_\Psi(\Lambda)$ to all of $\overline{\mathcal{L}}_+$ by setting $J_\Psi(\Lambda) := J_\Psi^\infty(\Lambda)$ on $\partial\mathcal{L}_+$. Notice that, by (20), J_Ψ is finite and given by (16) on $\overline{\mathcal{L}}_+ \setminus \{0\}$, and it is $+\infty$ in $\Lambda = 0$.

Lemma 3. *Assume (4). Then*

$$\lim_{||\Lambda||\to+\infty} J_\Psi(\Lambda) = +\infty. \tag{21}$$

Proof. Recall that by (4), there exists $\Phi_1 \in \mathcal{S}_+(\mathbb{T})$ satisfying (5), and, consequently, $\mathrm{tr}\,(\Lambda) = \int G^*\Lambda G\Phi_1 > 0, \forall\Lambda \in \mathcal{L}_+$. Suppose Λ_k is a sequence of matrices in $\mathcal{L}_+$ such that $\lim_{k\to\infty} ||\Lambda_k|| = +\infty$. Define the normalized sequence $\Lambda_k^0 := \frac{\Lambda_k}{||\Lambda_k||}$ (of course, we can assume $\Lambda_k \neq 0, \forall k$). Since $\mathrm{tr}\,\Lambda_k^0 > 0$,

$$\eta := \liminf_{k\to+\infty} \mathrm{tr}\,\Lambda_k^0 \geq 0.$$

Consider a sub-sequence such that the limit of its trace is η. This subsequence contains a convergent sub-subsequence $\{\Lambda_{k_m}^0\}$ since Λ_k^0 belongs to the surface of the unit ball, which is compact. Let $\Lambda_\infty := \lim_{m\to\infty} \Lambda_{k_m}^0$. Since $G^*\Lambda_n^0 G > 0$ on $\mathbb{T}$, $G^*\Lambda_\infty G \geq 0$

on $\mathbb{T}$. Moreover, $\Lambda_\infty \in \operatorname{Range}\Gamma$, since $\operatorname{Range}\Gamma$ is finite-dimensional, and hence closed. This implies that $G^*\Lambda_\infty G$ cannot be identically zero. In fact, if so, $\Lambda_\infty \in \mathcal{L}_+ = \ker(\Xi) = \operatorname{Range}\Gamma^\perp$. Then $\Lambda_\infty \in \operatorname{Range}\Gamma \cap \operatorname{Range}\Gamma^\perp = \{0\}$, which is a contradiction since $\|\Lambda_\infty\| = 1$. Thus

$$\eta = \lim_{n\to\infty} \operatorname{tr}\Lambda_n^0 = \operatorname{tr}\Lambda_\infty = \int G^*\Lambda_\infty G\Phi_1 > 0 \tag{22}$$

Hence, there exists a K such that $\operatorname{tr}\Lambda_k^0 > \eta/2$ for all $k \geq K$. Finally, since $G^*\Lambda_k^0 G \leq G^*G$, we obtain:

$$\begin{aligned}
\liminf_{k\to\infty} J_\Psi(\Lambda_k) &= \liminf_{k\to\infty} ||\Lambda_k||\operatorname{tr}\Lambda_k^0 - \int \Psi \log ||\Lambda_k|| G^*\Lambda_k^0 G \\
&= \liminf_{k\to\infty} ||\Lambda_k||\operatorname{tr}\Lambda_k^0 - (\textstyle\int \Psi)\log ||\Lambda_k|| - \displaystyle\int \Psi \log G^*\Lambda_k^0 G \\
&\geq \liminf_{k\to\infty} ||\Lambda_k||\operatorname{tr}\Lambda_k^0 - (\textstyle\int \Psi)\log ||\Lambda_k|| - \displaystyle\int \Psi \log G^*G \\
&\geq \liminf_{k\to\infty} ||\Lambda_k||\frac{\eta}{2} - (\textstyle\int \Psi)\log ||\Lambda_k|| - \displaystyle\int \Psi \log G^*G \\
&= \liminf_{k\to\infty} \frac{\eta}{2}\left(||\Lambda_k|| - \frac{\int\Psi}{\eta/2}\log||\Lambda_k||\right) - \int \Psi \log G^*G \\
&= +\infty.
\end{aligned}$$

■

Theorem 2. *Assume that the feasibility condition (4) is satisfied. Then the problem of minimizing the functional* $J_\Psi(\Lambda) = \operatorname{tr}\Lambda - \int \Psi \log G^*\Lambda G$ *over* $\mathcal{L}_+$ *admits a unique solution* $\hat{\Lambda} \in \mathcal{L}_+$.

Proof. In view of Lemma 1, Lemma 2 and Lemma 3, the functional J_Ψ is inf-compact on the closed set $\overline{\mathcal{L}}_+$, and therefore it admits a minimum point $\hat{\Lambda}$ there. We show next that $\hat{\Lambda} \in \mathcal{L}_+$. Of course, $\hat{\Lambda}$ is not the zero matrix since $J_\Psi(0) = +\infty$. Let $0 \neq \overline{\Lambda} \in \partial\mathcal{L}_+$. By Lemma 2, $J_\Psi(\overline{\Lambda})$ is finite. Observe that, by (4), $I \in \mathcal{L}_+$. By convexity of $\overline{\mathcal{L}}_+$, it then follows that $\overline{\Lambda} + \epsilon(I - \overline{\Lambda}) \in \overline{\mathcal{L}}_+, \forall\epsilon \in [0,1]$. We compute the one-sided directional derivative or hemidifferential

$$\begin{aligned}
J'_{\Psi_+}(\overline{\Lambda}; I - \overline{\Lambda}) &:= \lim_{\epsilon \searrow 0}\left[\frac{J_\Psi(\overline{\Lambda} + \epsilon(I - \overline{\Lambda})) - J_\Psi(\overline{\Lambda})}{\epsilon}\right] \\
&= \operatorname{tr}(I - \overline{\Lambda}) + \int \Psi - \int \frac{G^*G\Psi}{G^*\overline{\Lambda}G} = -\infty.
\end{aligned} \tag{23}$$

Hence, $\overline{\Lambda}$ cannot be a minimum point. We conclude that $\hat{\Lambda} \in \mathcal{L}_+$. ■

6 A Descent Method for the Dual Problem

In general, the optimal solution of the dual problem needs to be computed numerically. This is a delicate problem because of the unboundeness of the gradient of J_Ψ at the boundary of $\mathcal{L}_+$, see (23). The approaches proposed in [30] and references therein involve some preliminary reparametrization of $\mathcal{L}_+$, which may imply loss of global convexity.

In [41], a different matricial iterative method was proposed that appears to be very fast and numerically robust. This method does not restrict the search of $\hat{\Lambda}$ to $\mathcal{L}_+$ and indeed it normally converges to a $\hat{\Lambda} \notin \mathrm{Range}(\Gamma)$. We show below that this method may be viewed as a modified *gradient descent method* with fixed step size. This method is described as follows.

Let

$$\mathcal{M} := \{M \in \mathcal{L}'_+ :\ 0 \leq M \leq I,\ \mathrm{tr}\,[M] = 1\}, \tag{24}$$

$$\mathcal{M}_+ := \{M \in \mathcal{M} :\ M > 0\}. \tag{25}$$

For $M \in \mathcal{M}$, define the map Θ by

$$\Theta(M) := \int M^{1/2} G \left[\frac{\Psi}{G^* M G}\right] G^* M^{1/2}. \tag{26}$$

Theorem 3. [41]. *The map Θ maps $\mathcal{M}$ into $\mathcal{M}$ and $\mathcal{M}_+$ into $\mathcal{M}_+$.*

Consider the following iterative algorithm.

Algorithm. Let $M_0 = \frac{1}{n} I$. Note that $M_0 \in \mathcal{M}_+$. Define the sequence $\{M_k\}_{k=0}^{\infty}$ by

$$M_{k+1} := \Theta(M_k). \tag{27}$$

Notice that, by Theorem 3, $M_k \in \mathcal{M}_+$ for all k. Moreover, since $M_k \in \mathcal{M}, \forall k$, the sequence is bounded. Hence it has at least one accumulation (limit) point in the closure $\overline{\mathcal{M}}$ of $\mathcal{M}$.

Theorem 4. *Suppose that the sequence $\{M_k\}_{k=0}^{\infty}$ has a limit $\hat{M} \in \mathcal{M}_+$. Then $\hat{M} \in \mathcal{L}'_+$ and satisfies (11), and therefore provides the optimal solution of the approximation problem via (12).*

Notice that even when the sequence generated by (27) converges to a singular matrix $\hat{M} \in \mathcal{M}$, it is still possible, though not guaranteed, that such a matrix solves the original problem. We next show that the algorithm may be viewed as a modified gradient descent method. To this aim, rewrite (27) as

$$M_{k+1} = M_k + M_k^{1/2} \left[\int \frac{G \Psi G^*}{G^* M_k G} - I\right] M_k^{1/2}. \tag{28}$$

Proposition 1. *Define*

$$\Delta M_k := M_k^{1/2} \left[\int \frac{G \Psi G^*}{G^* M_k G} - I\right] M_k^{1/2}, \tag{29}$$

so that (28) reads $M_{k+1} = M_k + \Delta M_k$. Then, ΔM_k is a descent direction *at M_k for J_Ψ.*

Proof. Let

$$\nabla J_\Psi(M_k) = I - \int \frac{G\Psi G^*}{G^* M_k G}$$

denote the "gradient" of J_Ψ at M_k. Then,

$$< \nabla J_\Psi(M_k), \Delta M_k >= \text{tr}\ (\nabla J_\Psi(M_k)\Delta M_k) = -\text{tr}\ \left(M_k^{1/4}\nabla J_\Psi(M_k)M_k^{1/4}\right)^2.$$

By Theorem3, $M_k > 0$, for all k. It follows that $\text{tr}\ (\nabla J_\Psi(M_k)\Delta M_k) < 0$, unless $\nabla J_\Psi(M_k) = 0$ in which case M_k is a fixed point of the iteration which solves the dual problem by Theorem(4). ■

One could implement the matricial iteration as

$$M_{k+1} = M_k + \alpha_k \Delta M_k, \tag{30}$$

where $0 < \alpha_k \leq 1$ is determined through backstepping, see e.g. [5]. Our extensive simulation (see e.g. [41]), however, shows that convergence in fact occurs with $\alpha_k \equiv 1$! Indeed, the algorithm appears to perform numerically very well. In fact, at each step the integral (26) may be computed very precisely and efficiently via a spectral factorization technique that only requires to solve an algebraic Riccati equation and a Lyapunov equation, both of dimension n. We have performed an extensive number of simulations where the sequence generated by (27) never failed to converge. In a very small number of cases, we have observed convergence toward a singular matrix which, however, satisfied (11), and therefore provided the optimal solution of the approximation problem.

References

1. H. Akaike, *Markovian representation of stochastic processes by canonical variables*, SIAM J. Contr. vol. **13**, pp. 162-173, 1975
2. H. Akaike, *Stochastic theory of minimal realization*, IEEE Trans. Aut. Contr. vol. **AC-19**, pp. 667-674, 1974
3. A. Barron, *Entropy and the central limit theorem*, Ann. Probab. vol. **14**, pp. 336-342, 1986.
4. A.Blomqvist, A.Lindquist and R.Nagamune, *Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimizaiton approach*, IEEE Trans. Aut. Control vol. **48**, pp. 2172-2190, 2003.
5. S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
6. C. I. Byrnes, T. Georgiou, and A. Lindquist, *A new approach to spectral estimation: A tunable high-resolution spectral estimator*, IEEE Trans. Sig. Proc. vol. **49**, pp. 3189-3205, 2000.
7. C. I. Byrnes, T. Georgiou, and A. Lindquist, *A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint*, IEEE Trans. Aut. Control vol. **46**, pp. 822-839, 2001.
8. C. I. Byrnes, T. Georgiou, A. Lindquist and A. Megretski, *Generalized interpolation in H-infinity with a complexity constraint*, Trans. American Math. Society vol. **358(3)**, pp. 965-987, 2006 (electronically published on December 9, 2004).

9. C. I. Byrnes, S. Gusev, and A. Lindquist, *A convex optimization approach to the rational covariance extension problem*, SIAM J. Control and Opimization vol. **37**, pp. 211-229, 1999.
10. C. I. Byrnes, S. Gusev, and A. Lindquist, *From finite covariance windows to modeling filters: A convex optimization approach*, SIAM Review vol. **43**, pp. 645-675, 2001.
11. C. I. Byrnes and A. Lindquist, *The generalized moment problem with complexity constraint*, Integral Equations and Operator Theory vol. **56(2)**, pp. 163-180, 2006 (published online March 29, 2006).
12. E. Carlen and A. Soffer, *Entropy production by convolution and central limit theorems with strong rate information*, Comm. Math. Phys. vol. **140**, pp. 339-371, 1991.
13. T. M. Cover and J. A. Thomas, *Information Theory*, Wiley, New York, 1991.
14. H. Cramér, *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, 1946.
15. I. Csiszár, *Maxent, mathematics and information theory*, in Proc. 15th Inter. Workshop on Maximum Entropy and Bayesian Methods, K.M. Hanson and R.N. Silver eds., Kluver Academic, pp. 35-50, 1996.
16. J. C. Doyle, B. A. Francis and A. R. Tannenbaum, *Feedback Control Theory*, Macmillan, New York, 1992.
17. P. Enquist, *A homotopy approach to rational covariance extension with degree constraint*, Int. J. Appl. Math. and Comp. Sci. vol. **11**, pp. 1173-1201, 2001.
18. A. Ferrante, M. Pavon and F. Ramponi, *Constrained spectrum approximation in the Hellinger distance*, preprint Oct. 2006. To appear in Proc. of ECC07 Conf. 2007.
19. A. Ferrante, M. Pavon and F. Ramponi, *Hellinger vs. Kullback-Leibler multivariable spectrum approximation*, submitted. 2007.
20. H. Föllmer, *Random fields and diffusion processes*, in École d'Été de Probabilités de Saint-Flour XV-XVII, edited by P. L. Hennequin, Lecture Notes in Mathematics, Springer-Verlag, New York, vol. **1362**, pp. 102-203, 1988.
21. T. Georgiou, *Realization of power spectra from partial covariance sequences*, IEEE Trans. on Acoustics, Speech, and Signal Processing vol. **35**, pp. 438-449, 1987.
22. T. Georgiou, *The interpolation problem with a degree constraint*, IEEE Trans. on Aut. Control vol. **44**, pp. 631-635, 1999.
23. T. Georgiou, *Spectral estimation by selective harmonic amplification*, IEEE Trans. Aut. Control vol. **46**, pp. 29-42, 2001.
24. T. Georgiou, *The structure of state covariances and its relation to the power spectrum of the input*, IEEE Trans. Aut. Control vol. **47**, pp. 1056-1066, 2002.
25. T. Georgiou, *Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization*, IEEE Trans. Aut. Control vol. **47**, pp. 1811-1823, 2002.
26. T. Georgiou, *Solution of the general moment problem via a one-parameter imbedding*, IEEE Trans. Aut. Control vol. **50**, pp. 811-826, 2005.
27. T. Georgiou, *Relative entropy and the multivariable multidimensional moment problem*, IEEE Trans. Inform. Theory vol. **52**, pp. 1052-1066, 2006.
28. T. Georgiou, *Distances between power spectral densities*, arXiv e-print `math.OC/0607026`.
29. T. Georgiou, *An intrinsic metric for power spectral density functions*, arXiv e-print `math.OC/0608486`.
30. T. Georgiou and A. Lindquist, *Kullback-Leibler approximation of spectral density functions*, IEEE Trans. Inform. Theory vol. **49**, pp. 2910-2917, 2003.
31. E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, R.D. Rosenkranz ed., Dordrecht, 1983.
32. S. Kullback, *Information Theory and Statistics 2nd ed.*, Dover, Mineola NY, 1968.
33. A. Lindquist and G.Picci, On the stochastic realization problem, *SIAM J. Control and Optimization* **17** (1979), 365-389.

34. A. Lindquist and G. Picci, Forward and backward semimartingale models for Gaussian processes with stationary increments, *Stochastics* **15** (1985), 1-50.
35. A. Lindquist and G. Picci, Realization theory for multivariate stationary Gaussian processes, *SIAM J. Control and Optimization* **23** (1985), 809-857.
36. A. Lindquist and G. Picci, A geometric approach to modeling and estimation of linear stochastic systems, *J. Mathematical Systems, Estimation, and Control* **1** (1991), 241-333.
37. A. Lindquist and G. Picci, Geometric methods for state space identification, in *Identification, Adaptation, Learning: The Science of Learning Models from Data*, S. Bittanti and G. Picci (editors), Nato ASI Series (Series F, Vol 153), Springer, 1996, 1–69.
38. A. Lindquist and G. Picci, Canonical correlation analysis, approximate covariance extension, and identification of stationary time series, *Automatica* **32** (1996), 709–733.
39. H. P. McKean Jr., *Brownian motion with a several-dimensional time*, Th. Probab. Applic. vol. **8**, pp. 335-354, 1963
40. R. Nagamune, *A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint*, IEEE Trans. Aut. Control vol. **48**, pp. 113-117, 2003.
41. M. Pavon and A. Ferrante, *On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities*, IEEE Trans. Aut. Control vol. **51**, pp. 639-644, 2006.
42. M. Pavon and F. Ticozzi, *On entropy production for controlled Markovian evolution*, J. Math. Phys., vol. **47**, 063301, 2006.
43. G. Picci, *Stochastic realization of Gaussian processes*, Proc. IEEE vol. **64**, pp. 112-122, 1976.
44. G. Picci, *Some connections between the theory of sufficient statistics and the identifiability problem*, SIAM J. Appl. Math. vol. **33**, pp. 383-398, 1977.
45. G. Picci, *On the internal structure of finite-state stochastic processes*, in Recent developements in Variable Structure Systems, R. Mohler and A. Ruberti Eds. eds., Springer Lecture Notes in Economics and Mathematical Systems, vol. 162, pp. 288–304, 1978.
46. G. Ruckebush, *Representations markoviennes de processus gaussiens stationnaires*, Thèse 3ème cycle, Paris VI, 1975.
47. Yu. A. Rozanov, *Stationary Random Processes*, Holden-Day, San Francisco, 1967.
48. V. Vedral, *The role of relative entropy in quantum information theory*, Rev. Mod. Phys vol. **74**, pp. 197-, 2002.