STRUCTURED COVARIANCE ESTIMATION IN HIGH RESOLUTION SPECTRAL ANALYSIS*

A. FERRANTE†, M. PAVON ‡, AND M. ZORZI§

Abstract. We deal with the problem of estimating the output covariance of a linear filter fed by a stationary process. The estimated covariance must be positive semi-definite and must be in a prescribed vector space. It turns out that the problem is nontrivial when imposing both constraints. In this paper, we present and compare different strategies available in literature.

Key words. Covariance matrix estimation, convex optimization, maximum entropy, Toeplitz matrices.

AMS subject classifications. 62G05, 93E10, 62J10

1. Introduction.

Let us consider a stationary time series $w = \{w_k\}_{k=-\infty}^{\infty}$ obtained as the output of a given finite dimensional linear filter $G(z) = (zI - A)^{-1}B$ fed by a stationary process $y = \{y_k\}_{k=-\infty}^{\infty}$. The corresponding output state covariance matrix $\Sigma = E[w_k w_k^*]$ ($*$ denotes transposition plus conjugation) belongs to a vector space $\mathcal{V}$ that can be described by the parameters of the linear filter $G(z)$, [11], [9]. Given a finite length-collection of sample data $y_1 \dots y_N$, an estimate of Σ is represented by the sample covariance $\hat{\Sigma}_C := \sum_{k=1}^{N} w_k w_k^*$ which is a positive semi-definite matrix. This estimate, however, does not belong in general to the prescribed vector space. Then, one could project $\hat{\Sigma}_C$ onto the vector space $\mathcal{V}$. The projected matrix, however, is not guaranteed to be positive semi-definite. Indeed, when N (the length of the data record) is small it often fails to be positive semi-definite.

The importance of the estimation of covariance matrices belonging to the vector space induced by $G(z)$ is due to the development of a family of spectral estimation methods introduced by Byrnes, Georgiou and Lindquist in [3], and [4], and further developed and modified in [10, 7, 6]. These methods, for which $y_1, \dots, y_N$ and $G(z)$ are the given data, are based on a moment problem that requires an estimate of the covariance matrix of the output w. The first of these spectral estimation methods was called "THREE", [3]: we shall thus refer to these methods as "THREE-like".

In this paper we characterize the vector space associated to a general filter of the form $G(z) = C(zI - A)^{-1}B + D$ thus extending part of the results presented in [11]. Then, we introduce three different approaches to estimate the output covariance matrix, compatible with the structure of the given filter $G(z)$, based on the knowledge of the input sample data $y_1, \dots, y_N$. The first method is a simple procedure which projects the sample covariance $\hat{\Sigma}_C$ onto the specific vector space. If the projected estimate fails to be positive semi-definite then it is "adjusted" in some specific way. Then, we introduce a maximum entropy approach, [8], which leads to a positive definite estimate $\hat{\Sigma}$ consistent with the filter structure. Notice that this technique requires that the

*Work partially supported by the Italian Ministry for Education and Research (MIUR) under PRIN grant n. 20085FFJ2Z "New Algorithms and Applications of System Identification and Adaptive Control".

†is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy augusto@dei.unipd.it.

‡is with the Dipartimento di Matematica Pura ed Applicata, Università di Padova, via Trieste 63, 35131 Padova, Italy pavon@math.unipd.it.

§is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy zorzimat@dei.unipd.it.

sample covariance $\hat{\Sigma}_C$ is strictly positive definite (which is a very mild assumption) and that the filter's output and state coincide. On the other hand, these techniques do not exploit the knowledge of $y_1, \ldots, y_N$. The third method is an extension, [16], of the *biased spectral estimator* [15] employed for estimating a covariance matrix of order M with *Toeplitz* structure. This third method hinges on the characterization of Σ in terms of the filter $G(z)$ and the covariance lags sequence of the input process y. Thus, given an estimate of the covariance lags sequence of the input process, we can compute a positive semi-definite estimate $\hat{\Sigma}$ consistent with the structure imposed by the filter. Finally, we compare the performances of the methods afore cited.

2. Problem formulation. Consider a linear filter

$$\begin{aligned} x_{k+1} &= Ax_k + By_k \\ w_k &= Cx_k + Dy_k, \quad k \in \mathbb{Z}, \end{aligned} \tag{2.1}$$

where $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, C \in \mathbb{C}^{p\times n}, D \in \mathbb{C}^{p\times m}$ and A has all its eigenvalues in the open unit disk. The input process y is $\mathbb{C}^m$-valued, wide sense stationary and purely nondeterministic. As mentioned in the Introduction, $\Sigma = \Sigma^* \geq 0$ denotes the covariance matrix of the (stationary) output process w and we denote by

$$G(z) = C(zI - A)^{-1}B + D \tag{2.2}$$

the filter transfer function. Let $\mathcal{H}_m$ be the m^2-dimensional, *real* vector space of Hermitian matrices of dimension $m \times m$, $\mathcal{H}_{m,+}$ be the intersection between $\mathcal{H}_m$ and the open cone of positive definite matrices and $\overline{\mathcal{H}}_{m,+}$ its closure. We denote by $C(\mathbb{T}, \mathcal{H}_m)$ the family $\mathcal{H}_m$-valued, continuous functions on the unit circle $\mathbb{T}$. Consider now the linear operator

$$\Gamma : C(\mathbb{T}, \mathcal{H}_m) \to \mathcal{H}_p, \quad \Psi \mapsto \int G\Psi G^*, \tag{2.3}$$

where integration takes place on $\mathbb{T}$ with respect to the normalized Lebesgue measure $d\vartheta/2\pi$. It follows that Σ belongs to the linear space

$$\operatorname{Range}\Gamma := \{M \in \mathcal{H}_p | \; \exists \Psi \in C(\mathbb{T}, \mathcal{H}_m) \text{ such that } \int G\Psi G^* = M\}. \tag{2.4}$$

Suppose now that A, B, C, D are known and a sample data $\{y_k\}_{i=1}^N$ is given. We want to compute an estimate $\hat{\Sigma}$ of Σ such that $\hat{\Sigma} \in \overline{[\operatorname{Range}\Gamma]}_+$, where $\overline{[\operatorname{Range}\Gamma]}_+$ denotes the closure of $[\operatorname{Range}\Gamma]_+ := \operatorname{Range}\Gamma \cap \mathcal{H}_{p,+}$. If we feed $G(z)$ with the data $\{y_k\}_{k=1}^N$ and collect the output data $\{w_k\}_{k=1}^N$, an estimate of Σ is given by the sample covariance $\hat{\Sigma}_C := \frac{1}{N}\sum_{k=1}^N w_k w_k^* \geq 0$. This estimate usually fails to belong to $\operatorname{Range}\Gamma$ which is typically a low-dimensional vector space in $\mathcal{H}_p$, see below. One could project $\hat{\Sigma}_C$ onto $\operatorname{Range}\Gamma$ obtaining a new Hermitian matrix $\hat{\Sigma}_\Gamma$. This matrix $\hat{\Sigma}_\Gamma$, however, may be indefinite and this is particularly likely when N is not large. In addition, the computation of a basis for $\operatorname{Range}\Gamma$ is not trivial.

3. Characterization of $\operatorname{Range}\Gamma$**.** Here we present some results concerning the range of the operator Γ. Most of proofs are omitted or only sketched: we refer to [8],[16] for all the details. We start by considering a particular, yet very relevant, situation. We will later deal with the general case.

3.1. State covariance matrices. We restrict attention to the case when $C = I_n$ and $D = 0_{n\times m}$, with $m < n = p$, so that Σ is a state covariance matrix. Under the additional assumptions that (A, B) is a reachable pair and B has full column rank, it was shown in [9], [11] (see also [14]), that an $n \times n$ matrix $M \in \mathcal{H}_n$ belongs to $\operatorname{Range}\Gamma$ if and only if there exists $H \in \mathbb{C}^{m\times n}$ such that

$$M - AMA^* = BH + H^*B^*. \tag{3.1}$$

Proposition 3.1. $\operatorname{Range}\Gamma$ *has* real dimension *equal to* $m(2n-m)$. *Moreover* $M \in \operatorname{Range}\Gamma$ *if and only if the following condition holds*

$$\Pi_B^\perp (M - AMA^*)\Pi_B^\perp = 0, \tag{3.2}$$

where we denote by $\Pi_B^\perp := I - B(B^*B)^{-1}B^*$ *the orthogonal projection onto* $[\operatorname{Range} B]^\perp$.

Proof. We only show the necessity of condition (3.2). We know that there exists $H \in \mathbb{C}^{m\times n}$ such that

$$M - AMA^* = BH + H^*B^*. \tag{3.3}$$

Pre and post-multiplying this relation by $\Pi_B^\perp$, we obtain

$$\Pi_B^\perp (M - AMA^*)\Pi_B^\perp = \Pi_B^\perp (BH + H^*B^*)\Pi_B^\perp = \Pi_B^\perp BH\Pi_B^\perp + [\Pi_B^\perp BH\Pi_B^\perp]^* = 0.$$

■

Then, it is possible to relax the reachability assumption.

Theorem 3.2. *Consider an* (A, B) *pair with* B *full column rank. Let* $T \in \mathbb{C}^{n\times n}$ *be a state space transformation such that the pair* $(T^{-1}AT, T^{-1}B)$ *is in standard reachability form. Let* l *be the dimension of the reachable subspace. Assume* $l > m$. *Then,* $\operatorname{Range}\Gamma$ *has real dimension equal to* $m(2l - m)$ *and* $M \in \operatorname{Range}\Gamma$ *if and only if there exists* $H_1 \in \mathbb{C}^{m\times l}$ *such that*

$$M - AMA^* = B\begin{bmatrix} H_1 & 0 \end{bmatrix} T^* + T\begin{bmatrix} H_1^* \\ 0 \end{bmatrix} B^*. \tag{3.4}$$

Note that the above theorem enables us to easily compute a basis for $\operatorname{Range}\Gamma$ also when the pair (A, B) is not reachable.

3.2. Characterization of $\operatorname{Range}\Gamma$ **in the general case.** We now consider a general linear filter $G(z) = C(zI - A)^{-1}B + D$ and the corresponding linear operator Γ defined in (2.3). Moreover, we define the linear operator

$$\Lambda : C(\mathbb{T}, \mathcal{H}_m) \to \mathcal{H}_{n+p}, \quad \Psi \mapsto \int L\Psi L^* \tag{3.5}$$

where

$$L(z) := \left(zI - \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} G_S(z) \\ z^{-1}G(z) \end{bmatrix} \tag{3.6}$$

and $G_S(z) = (zI - A)^{-1}B$.

Theorem 3.3. $M \in \operatorname{Range}\Gamma$ *if and only if there exist* $P \in \mathbb{C}^{n\times n}$ *and* $Q \in \mathbb{C}^{n\times p}$ *such that*

$$X := \begin{bmatrix} P & Q \\ Q^* & M \end{bmatrix} \in \operatorname{Range}\Lambda. \tag{3.7}$$

Note that, $L(z)$ satisfies the hypothesis of Theorem 3.2. Accordingly, we can compute a basis for $\operatorname{Range}\Lambda$.

4. State covariance matrix estimation. We start by considering the important special case where the output of the filter $G(z)$ coincides with its state, i.e. $C = I$, $D = 0$ and $n = p$. It is moreover assumed that the pair (A, B) is reachable and the spectral density of y is coercive (so that the to-be-estimated covariance Σ is necessarily strictly positive definite) and that the matrix B is full column rank. Accordingly, we want to compute an estimate $\hat{\Sigma} \in [\mathrm{Range}\,\Gamma]_+$ of $\Sigma \in \mathcal{H}_{n,+}$ from $\hat{\Sigma}_C \in \mathcal{H}_{n,+}$.

4.1. Projection method. This approach, [13, Section 8], [8], consists in projecting $\hat{\Sigma}_C$ onto $\mathrm{Range}\,\Gamma$. Denote by $\hat{\Sigma}_\Gamma$ the projected matrix. If $\hat{\Sigma}_\Gamma$ is not positive definite, it may be further adjusted adding a matrix of the form $\varepsilon\Sigma_+$ with $\Sigma_+ \in [\mathrm{Range}\,\Gamma]_+$ and $\varepsilon > 0$ so large that $\hat{\Sigma}_{PJ} := \hat{\Sigma}_\Gamma + \varepsilon\Sigma_+ > 0$. In this way, $\hat{\Sigma}_{PJ}$ is a positive definite estimate belonging to $\mathrm{Range}\,\Gamma$ of the true state covariance Σ. Note that (under the assumptions (A, B) reachable pair and B full column rank) a basis for $\mathrm{Range}\,\Gamma$ can be easily computed from (3.1). A positive definite matrix $\Sigma_+ \in \mathrm{Range}\,\Gamma$ indeed exists and can be easily computed as follows. Set $H_+ := \frac{1}{2}B^*$ and consider the equation

$$\Sigma_+ - A\Sigma_+A^* = BH_+ + H_+^*B^* = BB^*. \tag{4.1}$$

Since (A, B) is reachable and A is a stable matrix, we have that (4.1) admits a unique solution Σ_+ and such a solution is indeed positive definite. In view of (3.1), Σ_+ also belongs to $\mathrm{Range}\,\Gamma$.

4.2. Maximum entropy method. Recall that a fundamental (pseudo-)distance in mathematical statistics is the *information divergence* (Kullback-Leibler index, relative entropy), [5]. For two Gaussian distributions p_Σ, p_Ω on $\mathbb{R}^n$ with zero mean and covariance matrices $\Sigma \in \mathcal{H}_{n,+}$ and $\Omega \in \mathcal{H}_{n,+}$, respectively, it is given by

$$\mathbb{D}(p_\Sigma \| p_\Omega) := \frac{1}{2}\left[\log\det(\Sigma^{-1}\Omega) + \mathrm{tr}(\Omega^{-1}\Sigma) - n\right]. \tag{4.2}$$

Notice that the right-hand side of (4.2) provides a natural pseudo-distance, denoted henceforth by $\mathbb{D}(\Sigma\|\Omega)$, on the space $\mathcal{H}_{n,+}$. In [2, p.963], the above index is minimized with respect to the second argument obtaining a maximum-likelihood approach. Unfortunately, the solution can be computed only for special classes of problems. This fact leads us to consider the following problem: Given $\hat{\Sigma}_C \in \mathcal{H}_{n,+}$ and $G(z) = (zI - A)^{-1}B$, minimize $\mathbb{D}(\Sigma\|\hat{\Sigma}_C)$ over $\Sigma \in [\mathrm{Range}\,\Gamma]_+$. The solution provides the required estimate denoted as $\hat{\Sigma}_{ME}$.

In view of Proposition 3.1, our problem can be reformulated as follows: Find $\Sigma \in \mathcal{H}_{n,+}$ minimizing $\mathbb{D}(\Sigma\|\hat{\Sigma}_C)$ subject to the *linear* constraint $\Pi_B^\perp(\Sigma - A\Sigma A^*)\Pi_B^\perp = 0$. Thus, the problem resembles a most standard maximum entropy (or, equivalently, minimum relative entropy) problem [12], [5]. The Lagrangian function is

$$\begin{aligned} L(\Sigma, \Lambda) &= \mathbb{D}(\Sigma\|\hat{\Sigma}_C) + \mathrm{tr}\left[\Lambda\Pi_B^\perp(\Sigma - A\Sigma A^*)\Pi_B^\perp\right] \\ &= -\frac{1}{2}\log\det\Sigma + \frac{1}{2}\log\det\hat{\Sigma}_C + \frac{1}{2}\mathrm{tr}(\hat{\Sigma}^{-1}\Sigma) - \frac{n}{2} + \mathrm{tr}\left[\Lambda\Pi_B^\perp(\Sigma - A\Sigma A^*)\Pi_B^\perp\right]. \end{aligned} \tag{4.3}$$

The corresponding unconstrained problem consists in minimizing $L(\Sigma, \Lambda)$ over $\Sigma \in \mathcal{H}_{n,+}$. Then, the first variation at $\Sigma > 0$ in direction $\delta\Sigma \in \mathcal{H}_n$ is given by

$$\delta L(\Sigma, \Lambda, \delta\Sigma) = \mathrm{tr}[(-\frac{1}{2}\Sigma^{-1} + \frac{1}{2}\hat{\Sigma}_C^{-1} + \Pi_B^\perp\Lambda\Pi_B^\perp - A^*\Pi_B^\perp\Lambda\Pi_B^\perp A)\delta\Sigma].$$

By annihilating the first variation for each $\delta\Sigma \in \mathcal{H}_n$, we get that the form of the optimal solution is

$$\Sigma_\Lambda = \left(\hat{\Sigma}_C^{-1} + 2Q_\Lambda\right)^{-1}, \tag{4.4}$$

where

$$Q_\Lambda := \Pi_B^\perp \Lambda \Pi_B^\perp - A^* \Pi_B^\perp \Lambda \Pi_B^\perp A \tag{4.5}$$

belongs to $\operatorname{Range}\Gamma^\perp$, [8], and $\Lambda \in \mathcal{H}_n$ must satisfy the inequality

$$\hat{\Sigma}_C^{-1} + 2\Pi_B^\perp \Lambda \Pi_B^\perp - 2A^* \Pi_B^\perp \Lambda \Pi_B^\perp A > 0. \tag{4.6}$$

Thus, the original problem is reduced to finding $\hat{\Lambda}$ satisfying (4.6) and such that $\Sigma_{\hat{\Lambda}}$ satisfies (3.2). It turns out that the dual problem is equivalent to minimize the following function

$$J(\Lambda) := -\frac{1}{2} \operatorname{tr} \log \left(\hat{\Sigma}_C^{-1} + 2Q_\Lambda\right) \tag{4.7}$$

over the set

$$\mathcal{L} := \{\Lambda \in \mathcal{H}_n \mid \hat{\Sigma}_C^{-1} + 2Q_\Lambda > 0 \text{ and } \Pi_B^\perp \Lambda \Pi_B^\perp \neq 0\}. \tag{4.8}$$

Theorem 4.1. *The dual functional (4.7) has a unique minimum point in $\mathcal{L}$.*

Proof. Since $\mathcal{L}$ is a bounded set and J is strictly convex over it, [8], we only need to show that J takes a minimum value on $\mathcal{L}$. First we observe that J is continuous on its domain. We now demonstrate that J is inf-compact, i.e., the image of $(-\infty, r]$ under the map J^{-1} is a compact set. It is then sufficient to apply Weierstrass theorem which states that a continuous function defined on a compact set admits a minimum. Indeed, observing that $J(0) = \frac{1}{2} \log\det \hat{\Sigma}_C$, we can restrict the search for a minimum point to the image of $(-\infty, \frac{1}{2} \log\det \hat{\Sigma}_C]$ under J^{-1}. Since $\mathcal{L}$ is a bounded set, it is sufficient to show that

$$\lim_{\Lambda \to \partial\mathcal{L}} J(\Lambda) = +\infty. \tag{4.9}$$

Note that $\partial\mathcal{L}$ is the set of $\Lambda \in \mathcal{H}_n$, $\Pi_B^\perp \Lambda \Pi_B^\perp \neq 0$ for which $\hat{\Sigma}^{-1} + 2Q_\Lambda$ is a singular positive semi-definite matrix. Thus, for $\Lambda \to \partial\mathcal{L}$ all its eigenvalues remain bounded and at least one of them tends to 0^+. We denote with $\lambda_1, \ldots, \lambda_n > 0$ the eigenvalues of $\hat{\Sigma}^{-1} + 2Q_\Lambda$ and, without loss generality, we suppose that, for $\Lambda \to \partial\mathcal{L}$, $\lambda_1 \to 0^+$. Hence

$$\lim_{\Lambda \to \partial\mathcal{L}} J(\Lambda) = \lim_{\lambda_1 \to 0^+} -\frac{1}{2} \log \prod_{i=1}^n \lambda_i = \lim_{\lambda_1 \to 0^+} \sum_{i=1}^n \log \frac{1}{\sqrt{\lambda_i}} = +\infty. \tag{4.10}$$

■

5. Covariance matrix estimation in the general framework. We now deal with the general problem of finding an estimate $\hat{\Sigma} \in \overline{[\operatorname{Range}\Gamma]}_+$ of $\Sigma \in \overline{\mathcal{H}}_{n,+}$ given the sample covariance $\hat{\Sigma}_C \in \overline{\mathcal{H}}_{n,+}$ and $G(z)$ as in (2.2).

5.1. Constrained covariance estimation method. This approach exploits the following result which characterizes the output covariance Σ in terms of the filter parameters and the covariance lags of y. For the proof we refer to [16].

THEOREM 5.1. *Let y and w be the input and output processes of the linear filter $G(z)$ as defined in (2.2). Then, the covariance matrix of w_k is given by*

$$\Sigma = CPC^* + CQD^* + DQ^*C^* + DR_0D^* \tag{5.1}$$

where

$$Q := \sum_{j=1}^{\infty} A^{j-1}BR_j^*, \quad R_j := E[y_{k+j}y_k^*] \tag{5.2}$$

and P is the (unique) solution of the Lyapunov equation

$$P - APA^* = AQB^* + BQ^*A^* + BR_0B^*. \tag{5.3}$$

This enables us to define an estimate $\hat{\Sigma}_{CL}$ of the output covariance depending on an estimate $\hat{R}_j$ of the input covariance lags and to characterize the key feature $\hat{\Sigma}_{CL} \in \operatorname{Range}\Gamma$ in terms of a suitable property of the $\hat{R}_j$'s as follows.

COROLLARY 5.2. *Let*

$$T_M(R) := \begin{bmatrix} R_0 & R_{-1} & & R_{-M} \\ R_1 & \ddots & \ddots & \\ & \ddots & \ddots & R_{-1} \\ R_M & & R_1 & R_0 \end{bmatrix} \geq 0 \tag{5.4}$$

define the block-Toeplitz covariance matrix of order M of the process with covariance sequence $\{R_j\}_{j=0}^{\infty}$. Let $\{\hat{R}_j\}_{j=0}^{\infty}$ be a sequence of $m \times m$ matrices such that $T_M(\hat{R}) \geq 0$ for each $M \in \mathbb{N}$. Define

$$\hat{\Sigma}_{CL} := C\hat{P}C^* + C\hat{Q}D^* + D\hat{Q}^*C^* + D\hat{R}_0D^* \tag{5.5}$$

where $\hat{Q} := \sum_{j=1}^{\infty} A^{j-1}B\hat{R}_j^*$ and $\hat{P}$ is the (unique) solution to the Lyapunov equation

$$\hat{P} - A\hat{P}A^* = A\hat{Q}B^* + B\hat{Q}^*A^* + B\hat{R}_0B^*. \tag{5.6}$$

Then, $\hat{\Sigma}_{CL} \in \overline{[\operatorname{Range}\Gamma]}_+$.

Proof. Since $T_M(\hat{R}) \geq 0$ for each $M \in \mathbb{N}$, there exists a wide sense stationary $\mathbb{C}^m$-valued process $\hat{y}$ with covariance lags sequence $\{\hat{R}_j\}_{j=0}^{\infty}$. If we feed the filter $G(z)$ with $\hat{y}$, we get a stationary output process $\hat{w}$. In view of Theorem 5.1, it follows that the covariance matrix of $\hat{w}$ is $\hat{\Sigma}_{CL} \in \overline{[\operatorname{Range}\Gamma]}_+$. ∎

In [16] it was shown that a possible strategy is given by the *Blackman-Tukey method* [1] with a rectangular lag window of width equal to L

$$\hat{R}_j = \begin{cases} \frac{1}{N}\sum_{k=1}^{N-j} y_{k+j}y_k^*, & 0 \leq j < L \\ 0_{m\times m}, & j \geq L. \end{cases} \tag{5.7}$$

and L is chosen in such way that $\|A^{L-1}\| < \varepsilon$, where ε is a threshold constant.

5.1.1. Projection method in the general case. We now show how to extend the projection method of Section 3 to the general setting considered in this paper. Let us first consider the situation where (A, B) may be non reachable (so that $\Sigma \geq 0$ may be singular) but still $C = I$ and $D = 0$. In view of Theorem 3.2, we can easily compute a basis for $\mathrm{Range}\,\Gamma$. Accordingly, we are able to compute the corresponding projected matrix $\hat{\Sigma}_\Gamma$ of $\hat{\Sigma}$. Here $\Sigma_+ \geq 0$ may be singular because we have removed the reachability condition. However, when $\hat{\Sigma}_\Gamma$ is indefinite, there always exists $\varepsilon > 0$ such that $\hat{\Sigma}_{PJ} := \hat{\Sigma}_\Gamma + \varepsilon\Sigma_+ \geq 0$ because the null space of Σ_+ coincides with the orthogonal complement of the reachable subspace of the pair (A, B). Then, in view of Theorem 3.3, we can employ the above procedure in order to estimate the output state covariance matrix of the filter $L(z)$ when is fed by the sample data $\{y_k\}_{k=1}^N$. Let $\hat{X}_{PJ} \in \overline{[\mathrm{Range}\,\Lambda]}_+$ denote the estimate, then $\hat{\Sigma}_{PJ} = \left[\begin{array}{cc} 0 & I_p \end{array}\right] \hat{X}_{PJ} \left[\begin{array}{c} 0 \\ I_p \end{array}\right]$.

6. Performance comparison. Here we sum up the comparison results among the constrained covariance estimation (CL), projection (PJ) and maximum entropy (ME) methods found once fixed $G(z)$ an the length N of the observed sequence. A detailed description can be found in [8],[16]. For each method we take into account the means μ_{CL}, μ_{PJ}, μ_{ME} and the variances σ^2_{CL}, σ^2_{PJ}, σ^2_{ME} of the corresponding relative error sequences $\{e_{CL}(j)\}$, $\{e_{PJ}(j)\}$, $\{e_{ME}(j)\}$ indexed by j. For each j, we choose a different sample path $y_1 \dots y_N$ in order to make the comparison reasonably independent of the specific data set. For example, the relative error at the j-th experiment for the CL method with estimate $\hat{\Sigma}_{CL}(j)$ is

$$e_{CL}(j) = \frac{\|\hat{\Sigma}_{CL}(j) - \Sigma\|}{\|\Sigma\|}. \tag{6.1}$$

It turns out that CL and ME methods largely outperform the PJ method. The heuristic reason follows: the projection of $\hat{X}_C$ (that is a perturbed version of the state covariance X) onto $\mathrm{Range}\,\Lambda$ yields a matrix $\hat{X}_\Lambda$ that, in many cases, in particular when N is small, fails to be positive definite (or even positive semi-definite). Moreover, when $\hat{X}_\Lambda$ is indefinite the projection method add to it the positive definite matrix $X_+ \in \mathrm{Range}\,\Lambda$. For each experiment, X_+ is the same. Accordingly the adjustment cannot be expected to provide a good estimate of $\hat{\Sigma}_{PJ}$. Then, CL and ME methods provide quite similar performances. The unique remarkable difference can be noted when Σ tend to be singular: Here the CL method outperforms in terms of computational burden the ME one. Finally, notice that ME method can be only used when Σ is a positive definite state covariance matrix.

7. Conclusion. In this paper, we have presented different approaches available in literature to estimate the covariance matrix Σ of the output processes of a given linear filter under the constraints of positivity and consistency with the structure imposed by the filter. The first one, PJ method, is a simple minded approach based on the projection principle. The second one is the CL method which hinges on an explicit representation of Σ in terms of the given filter and the covariance lags sequence of the input process. The last one, called ME method, is based on an optimization approach, which can be only employed to estimate positive definite state covariances. Finally, simulation evidence that CL and ME are preferable to the PJ method.

References.

[1] R. B. Blackman and J. W. Tukey. *The measurement of Power Spectra from the Point of View of Computation Engineering.* Dover, New York, 1958.

[2] J. P. Burg, D. G. Luenberger, and D. L. Wenger. Estimation structured covariance matrices. *Proceedings of the IEEE*, 70:963–974, 1982.

[3] C. I. Byrnes, T. Georgiou, and A. Lindquist. A new approach to spectral estimation: A tunable high-resolution spectral estimator. *IEEE Transaction on Signal Processing*, 49:3189–3205, 2000.

[4] C. I. Byrnes, T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Transaction on Automatic Control*, 46:822–839, 2001.

[5] T. M. Cover and J. A. Thomas. *Information Theory.* Wiley, New York, 1991.

[6] A. Ferrante, C. Masiero, and M. Pavon. Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. *IEEE Transaction on Automatic Control*, 2012 (to appear).

[7] A. Ferrante, M. Pavon, and F. Ramponi. Hellinger vs. Kullback-Leibler multivariable spectrum approximation. *IEEE Transaction on Automatic Control*, 53:954–967, 2008.

[8] A. Ferrante, M. Pavon, and M. Zorzi. A maximum entropy enhancement for a family of high-resolution spectral estimators. *IEEE Transaction on Automatic Control*, 57:318–329, 2012.

[9] T. Georgiou. Spectral estimation by selective harmonic amplification. *IEEE Transaction on Automatic Control*, 46:29–42, 2001.

[10] T. Georgiou. Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization. *IEEE Transaction on Automatic Control*, 47:1811–1823, 2002.

[11] T. Georgiou. The structure of state covariances and its relation to the power spectrum of the input. *IEEE Transaction on Automatic Control*, 47:1056–1066, 2002.

[12] E.T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, Sept. 1982.

[13] F. Ramponi, A. Ferrante, and M. Pavon. A globally convergent matricial algorithm for multivariate spectral estimation. *IEEE Transaction on Automatic Control*, 54(10):2376–2388, Oct. 2009.

[14] F. Ramponi, A. Ferrante, and M. Pavon. On the well-posedness of multivariate spectrum approximation and convergence of high-resolution spectral estimators. *Systems & Control Letters*, 59:167–172, 2010.

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

[16] M. Zorzi and A. Ferrante. On the estimation of structured covariance matrices. *Automatica*, Accepted, 2012.