Automatica 48 (2012) 2145–2151

Contents lists available at SciVerse ScienceDirect

Automatica

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

Brief paper

On the estimation of structured covariance matrices ☆

Mattia Zorzi [1], Augusto Ferrante

Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy

ARTICLE INFO

Article history:
Received 4 April 2011
Received in revised form
6 January 2012
Accepted 21 January 2012
Available online 30 June 2012

Keywords:
Covariance matrix estimation
Covariance lags sequence
Positive semi-definite matrix
Toeplitz matrices

ABSTRACT

This paper discusses a method for estimating the covariance matrix of a multivariate stationary process w generated as the output of a given linear filter fed by a stationary process y. The estimated covariance matrix must satisfy two constraints: it must be positive semi-definite and it must be consistent with the fact that w is the output of the given linear filter. It turns out that these constraints force the estimated covariance to lie in the intersection of a cone with a linear space. While imposing only the first of the two constraints is rather straightforward, guaranteeing that both are satisfied is a non-trivial issue to which quite a bit of attention has already been devoted in the literature. Our approach extends the method for estimating the *Toeplitz* covariance matrix of order M of a process y based on the *biased spectral estimator* (Stoica & Moses, 1997). This extension is based on the characterization of the output covariance matrix in terms of the filter parameters and the sequence of covariance lags of the input process.

After introducing our estimation method, we propose a comparison performance between this one and other methods proposed in the literature. Simulation results show that our approach constitutes a valid estimation procedure.

1. Introduction

In this paper, we consider the process $w = \{w_k\}_{k=-\infty}^{\infty}$ obtained as the output of a given stable rational filter $G(z)$ fed by a stationary process $y = \{y_k\}_{k=-\infty}^{\infty}$. We assume to observe a finite-length collection of sample data $y_1, \ldots, y_N$ of the stochastic process y. We want to compute an estimate $\hat{\Sigma}$ of the covariance $\Sigma := E[w_k w_k^*]$ in such a way that $\hat{\Sigma}$ is both positive semi-definite and consistent with the filter $G(z)$. Here $*$ denotes transposition plus conjugation. To analyze the features of this problem and to provide some motivations and applications, we discuss a very simple example. Let y be a real scalar second-order stationary process and let $G(z)$ be a bank of l delays:

$$G(z) := \begin{bmatrix} z^{-l} & z^{-l+1} & \cdots & z^{-1} \end{bmatrix}^\top. \tag{1}$$

In this case, the covariance matrix Σ of the output[2] w has the form of a symmetric *Toeplitz* matrix having the first l covariance lags of y on the first row:

$$\Sigma := \begin{bmatrix} r_0 & r_1 & \cdots & r_{l-1} \\ r_1 & r_0 & \ddots & r_{l-2} \\ \vdots & \ddots & \ddots & \ddots \\ r_{l-1} & \ddots & r_1 & r_0 \end{bmatrix}, \qquad r_h := E[y_{k+h} y_k^*]. \tag{2}$$

If we need to estimate Σ, it is natural to impose that the estimate $\hat{\Sigma}$ be positive semi-definite and have Toeplitz structure. On the one hand, one can consider the estimate $\hat{\Sigma}$ obtained by computing the sample covariance lags of y and constructing the corresponding Toeplitz matrix. This estimate, however, is *not* guaranteed to be positive semi-definite. On the other hand, one can compute the sample covariance $\hat{\Sigma}_C := \sum_{k=1}^{N} w_k w_k^*$ of the output process w. The latter is, by construction, positive semi-definite but is *not* guaranteed to be Toeplitz. Notice, in passing, that the orthogonal projection of this estimate onto the linear space of Toeplitz matrices is no longer guaranteed to be positive semi-definite. This problem, yet important, is very special due to the FIR structure of $G(z)$ in (1). In this case, it is well-known that the problem can be solved by computing, from $y_1, \ldots, y_N$, the estimates $\hat{r}_h$ of the r_h in (2), with the *biased correlogram spectral estimator* (Stoica & Moses, 1997). Alternatively, one can use a constrained convex optimization approach (Burg, Luenberger, & Wenger, 1982; Ferrante, Pavon, & Zorzi, 2012).

The estimation of positive semi-definite Toeplitz matrices is just an instance of a class of problems in digital signal processing

☆ 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". The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Er-Wei Bai under the direction of Editor Torsten Söderström.

E-mail addresses: zorzimat@dei.unipd.it (M. Zorzi), augusto@dei.unipd.it (A. Ferrante).

[1] Tel.: +39 049 8277702; fax: +39 049 8277799.

[2] In this case the output coincides with the state process.

0005-1098/$ – see front matter
doi:10.1016/j.automatica.2012.05.057

where the covariance matrix of the output process of a general linear filter has to be estimated with the knowledge of the input sample data. The importance of these problems is due to the development of a family of spectral estimation methods introduced by Byrnes, Georgiou and Lindquist in Byrnes, Georgiou, and Lindquist (2000), and Byrnes, Georgiou, and Lindquist (2001), and further developed and modified in Georgiou (2002a), Ferrante, Masiero, and Pavon (in press) and Ferrante, Pavon, and Ramponi (2008). These methods, for which $y_1, \ldots, y_N$ 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", Byrnes et al. (2000): we shall thus refer to these methods as "THREE-like".

For the special case of linear filters $G(z)$ whose output is the state of the filter, the problem of characterizing the output covariance Σ has been addressed by Georgiou in Georgiou (2001) and Georgiou (2002b). This characterization can be employed to estimate the state covariance by resorting to the maximum likelihood approach proposed in Burg et al. (1982) which, however, requires that the state covariance Σ and the sample covariance $\hat{\Sigma}_C$ are strictly positive definite. In Ferrante et al. (2012), a maximum entropy problem has been proposed that leads to a positive definite estimate $\hat{\Sigma}$ consistent with the filter structure. Notice that also this technique requires that the state covariance Σ and the sample covariance $\hat{\Sigma}_C$ are strictly positive definite and that the filter's output and state coincide. On the other hand, these techniques do not exploit the knowledge of $y_1, \ldots, y_N$ that, in the THREE-like methods, are the problem data.

The purpose of this paper is to introduce a new approach—based on the knowledge of the input sample data $y_1, \ldots, y_N$—to compute a positive semi-definite estimate $\hat{\Sigma}$ whose structure is consistent with an arbitrary, finite dimensional, stable, linear filter $G(z)$. Our method, which is an extension of the one for estimating the *Toeplitz* covariance matrix of order M of the process y based on the *biased spectral estimator* (Stoica & Moses, 1997), 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 an estimate $\hat{\Sigma}$ consistent with the structure imposed by the filter. It will be shown that if we consider the sample covariance lags used in the biased correlogram spectral estimator we can guarantee that $\hat{\Sigma} \geq 0$.

The paper is organized as follows. In the next section, we present a more precise formulation of the problem. In Section 3, the vector space containing the covariance matrices Σ is characterized in terms of the filter $G(z)$. Section 4 is devoted to introduce our approach based on the covariance lags. In Section 5, we briefly discuss other approaches available in the literature and their possible generalizations. Section 6 is devoted to simulations: we compare covariance matrices estimated by our method with the ones obtained using alternative approaches. In Section 7, we draw our conclusions.

2. Problem formulation

Consider a linear filter

$$x_{k+1} = Ax_k + By_k$$
$$w_k = Cx_k + Dy_k, \quad k \in \mathbb{Z}, \tag{3}$$

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

$$G(z) = C(zI - A)^{-1}B + D \tag{4}$$

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

$$\Gamma : C(\mathbb{T}, \mathfrak{H}_m) \to \mathfrak{H}_p, \quad \Psi \mapsto \int G\Psi G^*, \tag{5}$$

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

$$\text{Range}\,\Gamma := \left\{ M \in \mathfrak{H}_p \mid \exists \Psi \in C(\mathbb{T}, \mathfrak{H}_m) \text{ such that } \int G\Psi G^* = M \right\}. \tag{6}$$

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

$$\hat{\Sigma} \in [\text{Range}\,\Gamma]_+ := \text{Range}\,\Gamma \cap \mathfrak{H}_{p,+}. \tag{7}$$

If we feed $G(z)$ with the data $\{y_k\}_{k=1}^N$ and we 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, as it happened in the example discussed in the Introduction, normally fails to belong to Range Γ. In fact, Range Γ is a linear vector subspace usually strictly contained in $\mathfrak{H}_p$. One could project $\hat{\Sigma}_C$ onto Range Γ 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, when the linear filter $G(z)$ does not satisfy particular properties, the computation of a basis for Range Γ is not trivial.

3. Characterization of Range Γ

We start by considering a particular, yet very relevant, situation. We will later deal with the general case.

3.1. State covariance matrices

Next we restrict attention to the case when $C = I_n$ and $D = 0_{n\times m}$, with $m < n$, 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 Georgiou (2001) and Georgiou (2002b) (see also Ramponi, Ferrante, & Pavon, 2010), that an $n \times n$ matrix $M \in \mathfrak{H}_n$ belongs to Range Γ if and only if there exists $H \in \mathbb{C}^{m\times n}$ such that

$$M - AMA^* = BH + H^*B^*. \tag{8}$$

Moreover, it is possible to prove that Range Γ has *real dimension* equal to $m(2n - m)$, Ferrante et al. (2012).

We now want to relax the reachability assumption. To this end, we derive a preliminary result. Consider an (A, B) pair and the operator Γ corresponding to $G(z) = (zI - A)^{-1}B$. We perform a state space transformation induced by an invertible matrix $T \in \mathbb{C}^{n\times n}$,

$$\tilde{A} := T^{-1}AT, \qquad \tilde{B} := T^{-1}B. \tag{9}$$

We define the corresponding operator

$$\tilde{\Gamma} : C(\mathbb{T}, \mathfrak{H}_m) \to \mathfrak{H}_n, \qquad \Psi \mapsto \int \tilde{G}\Psi\tilde{G}^* \tag{10}$$

with $\tilde{G}(z) = (zI - \tilde{A})^{-1}\tilde{B} = T^{-1}G(z)$. Note that

$$\int G\Psi G^* = \int T\tilde{G}\Psi\tilde{G}^*T^*, \quad \forall \Psi \in C(\mathbb{T}, \mathfrak{H}_m). \tag{11}$$

Thus, Range Γ and Range $\tilde{\Gamma}$ are isomorphic vector spaces and

$$\tilde{M} \in \text{Range } \tilde{\Gamma} \Leftrightarrow T\tilde{M}T^* \in \text{Range } \Gamma. \quad (12)$$

Theorem 1. *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,* Range Γ *has real dimension equal to $m(2l - m)$ and $M \in$* Range Γ *if and only if there exists $H_1 \in \mathbb{C}^{m\times l}$ such that*

$$M - AMA^* = B\begin{bmatrix} H_1 & 0 \end{bmatrix} T^* + T\begin{bmatrix} H_1^* \\ 0 \end{bmatrix} B^*. \quad (13)$$

Proof. The proof is divided in three steps.

Step (1) By assumption, we have

$$\tilde{A} := T^{-1}AT = \begin{bmatrix} A_1 & A_{12} \\ 0 & A_2 \end{bmatrix}, \qquad \tilde{B} := T^{-1}B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \quad (14)$$

where $A_1 \in \mathbb{C}^{l\times l}, A_{12} \in \mathbb{C}^{l\times(n-l)}, A_2 \in \mathbb{C}^{(n-l)\times(n-l)}, B_1 \in \mathbb{C}^{l\times m}$, (A_1, B_1) reachable and B_1 full column rank. Then, it is easy to see that

$$\tilde{G}(z) = (zI - \tilde{A})^{-1}\tilde{B} = \begin{bmatrix} G_1(z) \\ 0 \end{bmatrix} \quad (15)$$

with $G_1(z) = (zI - A_1)^{-1}B_1$. Moreover, for each $\Psi \in C(\mathbb{T}, \mathfrak{H}_m)$ we have

$$\int \tilde{G}\Psi\tilde{G}^* = \begin{bmatrix} \int G_1\Psi G_1^* & 0 \\ 0 & 0 \end{bmatrix}. \quad (16)$$

Accordingly,

$$\text{Range } \tilde{\Gamma} = \left\{ \begin{bmatrix} M_1 & 0 \\ 0 & 0 \end{bmatrix} \text{ s.t. } M_1 \in \text{Range } \Gamma_1 \right\} \quad (17)$$

where

$$\Gamma_1 : C(\mathbb{T}, \mathfrak{H}_m) \to \mathfrak{H}_l, \qquad \Psi \mapsto \int G_1\Psi G_1^*. \quad (18)$$

It follows that Range Γ has the same dimension of Range Γ_1 and, since (A_1, B_1) is reachable and B_1 full column rank, as recalled before, such dimension is equal to $m(2l - m)$.

Step (2) Since (A_1, B_1) is reachable and B_1 full column rank, exploiting condition (8), we have that $\tilde{M} \in$ Range $\tilde{\Gamma}$ if and only if

$$\tilde{M} = \begin{bmatrix} M_1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad M_1 \in \mathbb{C}^{l\times l} \quad (19)$$

and there exists $H_1 \in \mathbb{C}^{m\times l}$ such that

$$M_1 - A_1M_1A_1^* = B_1H_1 + H_1^*B_1^*. \quad (20)$$

The above condition is equivalent to the existence of $H_1 \in \mathbb{C}^{m\times l}$ such that

$$\tilde{M} - \tilde{A}\tilde{M}\tilde{A}^* = \tilde{B}\begin{bmatrix} H_1 & 0 \end{bmatrix} + \begin{bmatrix} H_1^* \\ 0 \end{bmatrix}\tilde{B}^*$$
$$= \begin{bmatrix} B_1H_1 + H_1^*B_1^* & 0 \\ 0 & 0 \end{bmatrix}. \quad (21)$$

Here, we have exploited the fact that the (unique) solution $\tilde{M}$ of the Lyapunov equation (21) has the block-diagonal structure (19) with M_1 being the solution of (20).

Step (3) Pre and post multiplying (21) by T and T^*, respectively, we see that $\tilde{M} \in$ Range $\tilde{\Gamma}$ if and only if $\exists H_1 \in \mathbb{C}^{m\times l}$ such that $M := T\tilde{M}T^*$ satisfies (13). Exploiting (12) we obtain the statement. □

The previous theorem enables us to easily compute a basis for Range Γ also when the pair (A, B) is not reachable.

3.2. Characterization of Range Γ 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 (5). Moreover, we define the linear operator

$$\Lambda : C(\mathbb{T}, \mathfrak{H}_m) \to \mathfrak{H}_{n+p}, \qquad \Psi \mapsto \int L\Psi L^* \quad (22)$$

where

$$L(z) := \left(zI - \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} B \\ D \end{bmatrix} = \begin{bmatrix} G_S(z) \\ z^{-1}G(z) \end{bmatrix} \quad (23)$$

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

Theorem 2. *$M \in$ Range Γ if and only if there exist $P \in \mathbb{C}^{n\times n}$ and $Q \in \mathbb{C}^{n\times p}$ such that*

$$X := \begin{bmatrix} P & Q \\ Q^* & M \end{bmatrix} \in \text{Range } \Lambda. \quad (24)$$

Proof. Assume that $M \in$ Range Γ, then there exists $\Psi \in C(\mathbb{T}, \mathfrak{H}_m)$ such that $M = \int G\Psi G^*$. Define

$$P := \int G_S\Psi G_S^*, \qquad Q := \int e^{i\vartheta}G_S\Psi G^*. \quad (25)$$

It follows that

$$X := \begin{bmatrix} P & Q \\ Q^* & M \end{bmatrix} = \begin{bmatrix} \int G_S\Psi G_S^* & \int e^{i\vartheta}G_S\Psi G^* \\ \int e^{-i\vartheta}G\Psi G_S^* & \int G\Psi G^* \end{bmatrix}$$
$$= \int \begin{bmatrix} G_S \\ e^{-i\vartheta}G \end{bmatrix} \Psi \begin{bmatrix} G_S^* & e^{i\vartheta}G^* \end{bmatrix} = \int L\Psi L^*. \quad (26)$$

Accordingly $X \in$ Range Λ.

Conversely, assume that there exist P and Q such that (24) holds. Then there exists $\Psi \in C(\mathbb{T}, \mathfrak{H}_m)$ such that

$$X = \begin{bmatrix} P & Q \\ Q^* & M \end{bmatrix} = \int L\Psi L^*$$
$$= \begin{bmatrix} \int G_S\Psi G_S^* & \int e^{i\vartheta}G_S\Psi G^* \\ \int e^{-i\vartheta}G\Psi G_S^* & \int G\Psi G^* \end{bmatrix}. \quad (27)$$

Accordingly, $M = \int G\Psi G^*$, namely $M \in$ Range Γ. □

Note that, $L(z)$ satisfies the hypothesis of Theorem 1. Accordingly, we can compute a basis for Range Λ.

4. Constrained covariance estimation method

In this section, we first characterize the output covariance Σ in terms of the filter parameters and the covariance lags of y (Theorem 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 $\hat{\Sigma}_{CL} \in$ Range Γ in terms of a property of the $\hat{R}_j$'s (Corollary 4). Finally, we present a method to compute the $\hat{R}_j$'s guaranteeing $\hat{\Sigma}_{CL} \in [\text{Range } \Gamma]_+$.

Let $R_j := E[y_{k+j}y_k^*]$, $j \in \mathbb{Z}$, be the jth covariance lag of y. Notice that $R_j = R_{-j}^*$.

Theorem 3. *Let y and w be the input and output processes of the linear filter $G(z)$ as defined in (4). Then, the covariance matrix of w_k is given by*

$$\Sigma = CPC^* + CQD^* + DQ^*C^* + DR_0D^* \tag{28}$$

where

$$Q := \sum_{j=1}^{\infty} A^{j-1}BR_j^* \tag{29}$$

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

$$P - APA^* = AQB^* + BQ^*A^* + BR_0B^*. \tag{30}$$

Proof. From (3) we have

$$w_k w_k^* = Cx_kx_k^*C^* + Cx_ky_k^*D^* + Dy_kx_k^*C^* + Dy_ky_k^*D^*.$$

Taking expectations on both sides, we get (28), where $P := E[x_kx_k^*]$. Eq. (30) follows from Georgiou (2002b, Theorem 1). □

We now define the block-Toeplitz matrix

$$T_M(R) := \begin{bmatrix} R_0 & R_{-1} & & R_{-M} \\ R_1 & \ddots & \ddots & \\ & \ddots & \ddots & R_{-1} \\ R_M & & R_1 & R_0 \end{bmatrix} \tag{31}$$

as the covariance matrix of order M of the process y. Notice that, $T_M(R) \geq 0$ for each $M \in \mathbb{N}$.

Corollary 4. *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*

$$\hat{\Sigma}_{CL} := C\hat{P}C^* + C\hat{Q}D^* + D\hat{Q}^*C^* + D\hat{R}_0D^* \tag{32}$$

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

$$\hat{P} - A\hat{P}A^* = A\hat{Q}B^* + B\hat{Q}^*A^* + B\hat{R}_0B^*. \tag{33}$$

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

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 3, it follows that the covariance matrix of $\hat{w}$ is $\hat{\Sigma}_{CL} \in [\text{Range }\Gamma]_+$. □

Thus, once we have an estimate $\{\hat{R}_j\}_{j=0}^{\infty}$ of the covariance lags sequence of y satisfying $T_M(\hat{R}) \geq 0$ for each $M \in \mathbb{N}$, a positive semi-definite estimate $\hat{\Sigma}_{CL} \in$ Range Γ of the true covariance Σ is given by (32). It remains to choose a method to estimate $\{\hat{R}_j\}_{j=0}^{\infty}$ from the sample data $\{y_k\}_{k=1}^{N}$ in such a way that $T_M(\hat{R}) \geq 0$ for each $M \in \mathbb{N}$. We consider the correlogram spectral estimator, Stoica and Moses (1997), $\hat{\Phi} = \sum_{j=-\infty}^{\infty} \hat{R}_j e^{-i\vartheta j}$ where

$$\hat{R}_j = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{k=1}^{N-j} y_{k+j}y_k^*, & 0 \leq j < N \\ 0_{m\times m}, & j \geq N. \end{cases} \tag{34}$$

This method suffers from the drawback that the reliability of the estimate $\hat{R}_j$ decreases considerably as j grows, especially for relatively short time series, Kendall, Stuart, and Ord (1983). The corresponding estimated joint correlation $\hat{Q}$ is, however, a finite sum. Moreover, it is easy to see that $T_M(\hat{R}) = Y_MY_M^* \geq 0$ where $Y_M = \frac{1}{\sqrt{N}}\mathcal{C}$ with $\mathcal{C} \in \mathbb{C}^{mM\times(M-1+N)}$ being the left block-circulant (block Hankel) matrix, with m block rows, having $\begin{bmatrix} 0_{m\times 1} & \cdots & \cdots & 0_{m\times 1} & y_1 & \cdots & y_N \end{bmatrix}$ as the first block row. Notice that, in view of (28)–(30), the term A^{j-1} in $\hat{\Sigma}_{CL}$ acts as "reliability index" for the estimate $\hat{R}_j$. Indeed, due to the presence of the term A^{j-1}, the influence of $\hat{R}_j$ on $\hat{\Sigma}_{CL}$ decreases as j increases. Accordingly, we can truncate the covariance lags sequence in (34) to L

$$\hat{R}_j = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{k=1}^{N-j} y_{k+j}y_k^*, & 0 \leq j < L \\ 0_{m\times m}, & j \geq L. \end{cases} \tag{35}$$

L is chosen in such a way that $\|A^{L-1}\| < \varepsilon$, where ε is a threshold constant. Notice that (35) is the covariance lags sequence obtained by the *Blackman–Tukey method* (Blackman & Tukey, 1958) using a rectangular lag window of width equal to L. Thus, the corresponding block-Toeplitz matrix $T_M(\hat{R})$ is positive semi-definite for each M; see Stoica and Moses (1997). Hence, (35) is a natural choice for computing $\hat{\Sigma}_{CL}$.

The previous results suggest the following simple procedure, which we shall refer to as the *input covariance lags method*, to compute $\hat{\Sigma}_{CL}$ given the sample data $\{y_k\}_{k=1}^{N}$:

(1) Choose L such that $\|A^{L-1}\| < \varepsilon$.
(2) Compute

$$\hat{R}_0 = \sum_{k=1}^{N} y_ky_k^*, \qquad \hat{Q} = \frac{1}{N}\sum_{j=1}^{L-1}\sum_{k=1}^{N-j} A^{j-1}By_ky_{k+j}^*.$$

(3) Solve in $\hat{P}$ the Lyapunov equation (33).
(4) Compute the estimate $\hat{\Sigma}_{CL}$ of the true covariance Σ using (32).

5. Alternative methods

In this section, we discuss alternative methods for the structured covariance estimation problem.

5.1. State covariance matrix estimation

The problem addressed in this paper has so far been addressed in the literature under the following assumptions. The output of the filter $G(z)$ coincides with its state, i.e. $C = I$ and $D = 0$. Moreover, it was 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. Next we review these methods. In Section 5.2, we propose an extension of one of the methods to the general situation dealt with in this paper.

5.1.1. Projection method

In Ramponi, Ferrante, and Pavon (2009, Section 8) and Ferrante et al. (2012), a simple approach, called the projection method, was proposed to compute a positive definite estimate $\hat{\Sigma}_{PJ} \in$ Range Γ of Σ (that was assumed to be strictly positive definite) from the sample covariance $\hat{\Sigma}_C$. This method consists in projecting $\hat{\Sigma}_C$ onto Range Γ. Denote by $\hat{\Sigma}_\Gamma$ the projected matrix. If $\hat{\Sigma}_\Gamma$ is not positive definite, it may be further adjusted by adding a matrix of the form $\varepsilon\Sigma_+$ with $\Sigma_+ \in [\text{Range }\Gamma]_+$ 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 Range Γ of the true state covariance Σ. Note that (under the assumptions (A, B) reachable pair and B full column rank) a basis for Range Γ can be easily computed from (8). A positive definite matrix $\Sigma_+ \in$ Range Γ indeed exists and can be easily computed as follows. Set $H_+ := \frac{1}{2}B^*$ and consider the equation

$$\Sigma_+ - A\Sigma_+A^* = BH_+ + H_+^*B^* = BB^*. \tag{36}$$

Since (A, B) is reachable and A is a stable matrix, we have that (36) admits a unique solution Σ_+ and such a solution is indeed positive definite. In view of (8), Σ_+ also belongs to Range Γ.

5.1.2. Maximum entropy method

In Ferrante et al. (2012), a maximum entropy method was presented to determine a positive definite estimate $\hat{\Sigma}_{ME} \in \text{Range } \Gamma$ of the true state covariance $\Sigma > 0$. Consider the *information divergence* (Kullback–Leibler index, relative entropy, Cover & Thomas, 1991) between the two Gaussian distributions p_Σ, p_Ω on $\mathbb{R}^n$ with zero mean and covariance matrices $\Sigma > 0$ and $\Omega > 0$, respectively

$$\mathbb{D}(p_\Sigma \parallel p_\Omega) := \frac{1}{2}[\log\det(\Sigma^{-1}\Omega) + \text{tr}(\Omega^{-1}\Sigma) - n]. \tag{37}$$

Given the sample covariance $\hat{\Sigma}_C > 0$ the maximum entropy method computes $\hat{\Sigma}_{ME}$ by minimizing $\mathbb{D}(p_\Sigma \parallel p_{\hat{\Sigma}_C})$ with respect to Σ over the set $\{M = M^* > 0 \mid M \in \text{Range } \Gamma\}$. As observed in Burg et al. (1982, p. 963), $\mathbb{D}(\cdot \parallel \cdot)$ "really comes from maximum likelihood considerations and thus should, in some sense, gives us a reasonable answer, even if the process is not Gaussian and the vector samples are not independent". The optimal solution $\hat{\Sigma}_{ME}$ always exists and it is unique. Moreover, the maximum entropy method attains better performances than the ones obtained by the projection method; see Ferrante et al. (2012).

5.2. Projection method in the general case

We now show how to exploit the results of Section 3 to extend the projection method 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 1, we can easily compute a basis for Range Γ. 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).

In view of Theorem 2, we can now extend the projection method to the general case. Consider the linear filter $L(z)$ as in (23). Let v be the output process when $L(z)$ is fed by y

$$v_{k+1} = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} v_k + \begin{bmatrix} B \\ D \end{bmatrix} y_k, \quad k \in \mathbb{Z}. \tag{38}$$

Define then $X := E[v_k v_k^*]$ as the corresponding output covariance matrix. We are now ready to outline the generalization of the projection method. Let $\{v_k\}_{k=1}^N$ be the output data when $L(z)$ is fed with the sample data $\{y_k\}_{k=1}^N$. Compute then the sample matrix $\hat{X}_C := \frac{1}{N}\sum_{k=1}^N v_k v_k^*$. Notice that X is a state covariance matrix. Applying the projection method presented in Section 5.1.1, we obtain an estimate $\hat{X}_{PJ} \geq 0$ belonging to Range Λ. Finally, exploiting Theorem 2, we have

$$\hat{\Sigma}_{PJ} := \begin{bmatrix} 0 & I_p \end{bmatrix} \hat{X}_{PJ} \begin{bmatrix} 0 \\ I_p \end{bmatrix}. \tag{39}$$

6. Performance comparison

In this section, we want to test the method presented in Section 4 with the other methods sketched in the previous section. We use the following notation:

- The *CL method* to denote the input covariance lags method.
- The *PJ method* to denote the extended projection method presented in Section 5.2.
- The *ME method* to denote the maximum entropy method (only employed to estimate state covariance matrices).

For a fair interpretation of the comparison results, we hasten to point out that while other methods exploit only a finite sample of the output of the linear filter, our method uses only the corresponding sample of the input process. Notice that in the case of the estimate of state covariance matrices (and assuming the matrix B to be full column rank) the sample of the input process is easily obtainable from that of the output process and the converse is also true. Therefore, the available information is really the same for the three methods. On the contrary, for general filters the available information may be different. Notice also that in the applications related with THREE-like estimation methods the available data are a finite sample of the input process.

6.1. A performance comparison procedure

Suppose that we have a finite sequence $y_1, \ldots, y_N$ extracted from a sample path of a zero-mean, weakly stationary discrete-time process y. We want to compare the estimates $\hat{\Sigma}_{CL}, \hat{\Sigma}_{PJ}, \hat{\Sigma}_{ME}$ obtained by employing CL, PJ and ME methods, respectively. In order to make the comparison reasonably independent of the specific data set, we average over 500 experiments performed with sequences extracted from different sample paths. We are now ready to describe the comparison procedure:

- Fix the transfer function $G(z)$.
- At the jth experiment $G(z)$ is fed by the data $\{y_k^j\}_{k=1}^N$. From $\{y_k^j\}_{k=1}^N$ estimate $\hat{\Sigma}_{CL}(j)$, $\hat{\Sigma}_{PJ}(j)$ and $\hat{\Sigma}_{ME}(j)$ using CL, PJ and ME methods, respectively.
- Compute the relative error norm[3] between Σ and the estimate $\hat{\Sigma}_{CL}(j)$

$$e_{CL}(j) = \frac{\|\hat{\Sigma}_{CL}(j) - \Sigma\|}{\|\Sigma\|}. \tag{40}$$

  In a similar way, compute the relative error norms $e_{PJ}(j)$ and $e_{ME}(j)$ between Σ and the estimates $\hat{\Sigma}_{PJ}(j)$ and $\hat{\Sigma}_{ME}(j)$, respectively.
- Once completed the experiments, compute the means $\mu_{CL}, \mu_{PJ}, \mu_{ME}$ and the variances $\sigma_{CL}^2, \sigma_{PJ}^2, \sigma_{ME}^2$ of the corresponding sequences $\{e_{CL}(j)\}_{j=1}^{500}, \{e_{PJ}(j)\}_{j=1}^{500}, \{e_{ME}(j)\}_{j=1}^{500}$. For example, for the CL method:

$$\mu_{CL} = \frac{1}{500}\sum_{j=1}^{500} e_{CL}(j), \qquad \sigma_{CL}^2 = \frac{1}{500}\sum_{j=1}^{500}(e_{CL}(j) - \mu_{CL})^2.$$

- Count the number $\sharp F$ of times that the PJ method adjusts the estimate $\hat{X}_\Gamma$ by adding the quantity εX_+.

  Notice that the ME method can be only used when Σ is a state covariance matrix (and not in the general case). For the sake of comparison, we consider the parameters μ_i, σ_i^2 and $\sharp F$. Clearly, the smaller these parameters, the better estimation is expected.

6.2. Simulation results: the general case

We have considered a bivariate real process y with a high-order spectral density $\Phi(z)$ and a filter $G(z)$ with a 3-dimensional output with 4 poles equi-spaced on the circle of radius 0.8. The true covariance matrix Σ is positive definite with eigenvalues: $\lambda_1 = 3.12 \cdot 10^4, \lambda_2 = 1.15 \cdot 10^2, \lambda_3 = 3.33 \cdot 10^2$. The corresponding error means and variances for the PJ and CL methods are reported in Table 1 for different values of the length N of the

[3] Here the norm $\|\cdot\|$ is the spectral norm i.e. the matrix norm induced by the Euclidean norm in $\mathbb{C}^p$.

Table 1
Parameters $\mu_{CL}, \mu_{PJ}, \sigma^2_{CL}, \sigma^2_{PJ}, \sharp F$ for $G(z)$ considered in Section 6.2.

N	μ_{CL}	μ_{PJ}	σ^2_{CL}	σ^2_{PJ}	$\sharp F$
300	0.1360	0.4130	0.0089	2.0937	174
500	0.1016	0.2127	0.0045	0.5544	126
700	0.0865	0.1893	0.0031	0.8592	97

Table 2
Parameters $\mu_{CL}, \mu_{PJ}, \mu_{ME}, \sigma^2_{CL}, \sigma^2_{PJ}, \sigma^2_{ME}, \sharp F$ for $G(z)$ considered in Section 6.3.

N	μ_{CL}	μ_{PJ}	μ_{ME}	σ^2_{CL}	σ^2_{PJ}	σ^2_{ME}	$\sharp F$
300	0.18	0.81	0.18	0.018	2.65	0.02	73
500	0.16	0.47	0.15	0.013	1.37	0.013	37
700	0.13	0.29	0.13	0.001	0.74	0.009	18

observed data sequences $\{y_k\}_{k=1}^N$. It is clear that the CL method largely outperforms the PJ method. The heuristic reason follows. As noted in Ferrante et al. (2012), the projection of $\hat{X}_C$ (that is a perturbed version of the state covariance X) onto Range Λ 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). This, explains why the number of failures $\sharp F$ is significant. Moreover, when $\hat{X}_\Lambda$ is indefinite the projection method adds to it the positive definite matrix $X_+ \in$ Range Λ. For each experiment, X_+ is the same. In view of (39), the adjustment cannot be expected to provide a good estimate of $\hat{\Sigma}_{PJ}$. Note that μ_{PJ}, σ^2_{PJ} decrease as N increases: in fact, $\hat{X}_C \to X$ with probability one as $N \to \infty$. Notice that also μ_{CL} and σ^2_{CL} decrease as N grows. Indeed, each $\hat{R}_j$ approaches the true covariance lag R_j as $N \to \infty$. Accordingly $\hat{\Sigma}_{CL} \to \Sigma$. Moreover, each estimate $\hat{\Sigma}_{CL}$ is positive definite. We conclude that the CL method is remarkably preferable to the PJ method.

6.3. Simulation results: state covariance estimation

Consider $G(z)$ corresponding to $C = I_6,\ D = 0_{6\times 2}$,

$$A = \begin{bmatrix} 0.6 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0.7 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We choose the bivariate real process y with a coercive high-order spectral density $\Phi(z)$ considered in Ferrante et al. (2012, Section VII. C). The true covariance Σ is positive definite with eigenvalues: $\lambda_1 = 3.4 \cdot 10^{-3}, \lambda_2 = 1.69 \cdot 10^{-2}, \lambda_3 = 1.47, \lambda_4 = 2.92, \lambda_5 = 1.18 \cdot 10, \lambda_6 = 1.59 \cdot 10^2$. In Table 2, we present the results obtained for different lengths N of the observed sequences $\{y_k^j\}_{k=1}^N$. The CL and ME methods provide quite similar performances. The PJ method provides bad estimates when N is small. In this situation, the PJ method must adjust the projection $\hat{\Sigma}_C$ onto Range Γ in many experiments. Accordingly, its performance becomes remarkably poor with respect to the other methods when N is not large. Also in this case each estimate $\hat{\Sigma}_{CL}$ is positive definite.

Remark 5. As for the computational burden, the PJ method described in Section 5.2 normally compensates for the poor performances with a very high numerical efficiency. The ME and CL methods are very hardly comparable. In fact, the number of operations of the ME and CL methods is highly dependent on the problem's parameters. Moreover, the ME method is an optimization procedure whose computational burden is also dependent on the tolerance threshold fixed for the convergence of the algorithm. On the other hand, the CL does not require any optimization procedure. For example, in the cases illustrated above the two methods perform very similarly also with respect to the computational burden (while the PJ method is much faster). On the other hand, extensive simulation shows that the ME method presents numerical problems and leads to extremely slow convergence, if we consider a case when the state covariance Σ is close to singularity. Our method, on the contrary, does not require any optimization procedure and does not present any of these problems.

7. Conclusion

In this paper, we have proposed an efficient and natural approach 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. Our approach, called CL, hinges on an explicit representation of Σ in terms of the given filter and the covariance lags sequence of the input process. Not only the estimated matrix was shown to be positive semi-definite, but extensive simulation suggests also that the estimate is strictly positive definite with high probability when $\Sigma > 0$. We have also extended the PJ method to the general setting discussed in this paper and we have compared our CL method with this extended PJ method. It appears that, in several critical cases the CL method outperforms the other one.

In order to have a wider comparison, we have tested our method also against the maximum entropy approach (Section 5.1.2) in the restrictive framework where the latter method can be used. While the performances of these two methods are normally very similar, the CL method outperforms (in terms of computational burden) the ME method when Σ is close to singularity.

References

Blackman, R. B., & Tukey, J. W. (1958). *The measurement of power spectra from the point of view of computation engineering*. New York: Dover.

Burg, J. P., Luenberger, D. G., & Wenger, D. L. (1982). Estimation structured covariance matrices. *Proceedings of the IEEE*, *70*, 963–974.

Byrnes, C. I., Georgiou, T., & Lindquist, A. (2000). A new approach to spectral estimation: a tunable high-resolution spectral estimator. *IEEE Transactions on Signal Processing*, *49*, 3189–3205.

Byrnes, C. I., Georgiou, T., & Lindquist, A. (2001). A generalized entropy criterion for Nevanlinna–Pick interpolation with degree constraint. *IEEE Transactions on Automatic Control*, *46*, 822–839.

Cover, T. M., & Thomas, J. A. (1991). *Information theory*. New York: Wiley.

Ferrante, A., Masiero, C., & Pavon, M. (2012). Time and spectral domain relative entropy: a new approach to multivariate spectral estimation. In *IEEE Transactions on Automatic Control* (in press).

Ferrante, A., Pavon, M., & Ramponi, F. (2008). Hellinger vs. Kullback–Leibler multivariable spectrum approximation. *IEEE Transactions on Automatic Control*, *53*, 954–967.

Ferrante, A., Pavon, M., & Zorzi, M. (2012). A maximum entropy enhancement for a family of high-resolution spectral estimators. *IEEE Transactions on Automatic Control*, *57*, 318–329.

Georgiou, T. (2001). Spectral estimation by selective harmonic amplification. *IEEE Transactions on Automatic Control*, *46*, 29–42.

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

Georgiou, T. (2002b). The structure of state covariances and its relation to the power spectrum of the input. *IEEE Transactions on Automatic Control*, *47*, 1056–1066.

Kendall, M., Stuart, J., & Ord, J. (1983). *Advanced theory of statistics*: *Vol. 2* (4th ed.). New York: Macmillan.

Ramponi, F., Ferrante, A., & Pavon, M. (2009). A globally convergent matricial algorithm for multivariate spectral estimation. *IEEE Transactions on Automatic Control*, *54*, 2376–2388.

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

Stoica, P., & Moses, R. (1997). *Introduction to spectral analysis*. New York: Prentice Hall.

Mattia Zorzi was born in Padova, Italy, in January 1985. He received the B.S. degree (with high honors) in information engineering and the M.S. degree (with high honors) in automation engineering from University of Padova, Padova, Italy, in 2007 and 2009, respectively, where he is currently pursuing the Ph.D. degree.

His current research interests include spectral analysis and quantum systems estimation.

Augusto Ferrante was born in Piove di Sacco, Italy, on August 5, 1967. He received the "Laurea" degree (with high honors) in electrical engineering and the Ph.D. degree in control systems engineering from the University of Padova, Padova, Italy, in 1991 and 1995, respectively.

He has been on the faculty of the Colleges of Engineering of the University of Udine and of the "Politecnico di Milano". He is presently Professor in the Department of Information Engineering of the University of Padova. His research interests are in the areas of linear systems, spectral estimation, optimal control and optimal filtering, quantum control, and stochastic realization.