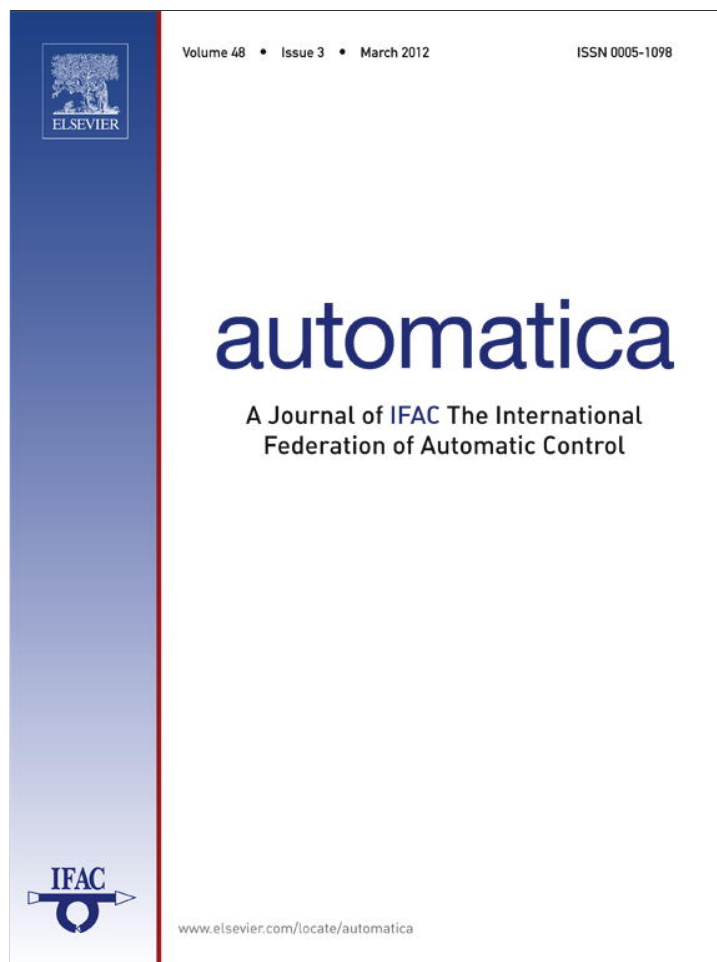




This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Brief paper

Consistency of subspace methods for signals with almost-periodic components[☆]Martina Favaro^a, Giorgio Picci^{b,1}^a Department of Electronics for Automation, University of Brescia, Italy^b Department of Information Engineering, University of Padova, Italy

ARTICLE INFO

Article history:

Received 29 June 2010

Received in revised form

11 May 2011

Accepted 7 August 2011

Available online 28 January 2012

Keywords:

Subspace identification

Frequency estimation

Statistical consistency

Almost periodic signals

Purely deterministic processes

ABSTRACT

It is sometimes claimed in the literature that subspace methods provide consistent estimates, also when the underlying observed signal has purely oscillatory modes (or the generating system has uncontrollable eigenvalues on the unit circle) but a formal proof of this assertion does not seem to exist. In this paper, we prove consistency of subspace methods with purely oscillatory modes. A well-known subspace identification procedure based on canonical correlation analysis and approximate partial realization is shown to be consistent under certain conditions on the purely deterministic part of the generating system. The algorithm uses a fixed finite regression horizon and the proof of consistency does not require that the regression horizon goes to infinity at a certain rate with the sample size N .

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

This paper deals with subspace identification of stationary processes with oscillatory components. At a first sight this problem may look like a minor generalization of a standard identification problem which has been exhaustively treated in the literature since the early 1990s. In reality, on one hand the problem encompasses harmonic retrieval; that is, estimation of the harmonic components of a stationary signal in additive noise, a problem of paramount importance in signal processing which, in the multichannel case, cannot be approached by the standard methods like Pisarenko, MUSIC, ESPRIT etc. It seems fair to say that the specialized literature on harmonic retrieval in the case of vector signals, when the additive noise is colored, is still far from offering satisfactory solutions. For this class of signals, on the other hand, subspace system identification appears as a natural choice.

However it is well-known that stationary random processes with periodic components are not ergodic. Non-ergodicity means in particular that the limit when the sample size goes to infinity of the process sample covariance is *sample dependent*. In particular,

[☆] This work was supported in part by the project *New methods and algorithms for Bayesian estimation, Identification and Adaptive Control* funded by the Italian Ministry of higher education (MIUR). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wolfgang Scherrer under the direction of Editor Torsten Söderström.

E-mail addresses: martina.favaro@ing.unibs.it, martinafavaro@yahoo.it (M. Favaro), picci@dei.unipd.it (G. Picci).

¹ Tel.: +39 049 8277705; fax: +39 049 8277614.

the limit sample covariance depends on the random amplitudes of its elementary oscillatory components; see, e.g. Söderström and Stoica (1989, pp. 105–109). On the other hand, the asymptotic statistical properties of subspace methods (and, more generally, of correlation-based methods) depend essentially on the limit sample covariances, which in the presence of oscillatory or quasi-periodic components are not equal to the ensemble averages; i.e., do not coincide with the true covariances. Since parameter estimation procedures based on correlation methods require solving linear relations involving estimated sample covariances, a natural question to ask is if the parameter estimates obtained by solving these linear equations are consistent. This is generally true for signals which are second-order ergodic but sample dependence casts doubts on the validity of standard asymptotic statistical properties, like consistency, of subspace methods in this setting. In particular legitimate doubts arise on the validity of the standard proofs of consistency of subspace methods for signals of this type.

Sections 4 and 5 deal with the question of asymptotically recovering the system parameters (modulo similarity) starting from finite data by a standard subspace algorithm, formulated as an approximate partial realization problem. This setting permits to prove almost sure consistency of the algorithm without having to estimate the transient estimation errors inherent in the truncated least-squares regression approach of Peternell (1995), Peternell, Deistler, and Scherrer (1995); Peternell, Scherrer, and Deistler (1996).

Consistency of subspace methods for purely non-deterministic signals (time series) has been proved earlier in the just cited references. However, to the best of the authors' knowledge, a proof

of consistency when there are quasi-periodic components due to uncontrollable eigenvalues on the unit circle, does not exist. The only paper which comes close in spirit to what concerns us here is [Bissacco, Chiuso, and Soatto \(2007\)](#). In this paper however consistency analysis had to be left out as being “beyond the scope of the paper”. Finally, note that processes described by systems whose eigenvalues of modulus one are reachable for the driving process noise, do not concern us here as these processes are actually non-stationary and do not contain almost-periodic oscillations.

2. Stationary processes with an almost-periodic component

All random variables/vectors, denoted by lowercase boldface characters, will have zero mean and finite second order moments. The symbol $\mathbb{E}$ denotes mathematical expectation. All random processes will be discrete time. It is a well-known fact that every vector-valued, say m -dimensional, second-order stationary process admits an orthogonal decomposition

$$\mathbf{y}(t) = \mathbf{y}_d(t) + \mathbf{y}_s(t), \quad t \in \mathbb{Z} \quad (2.1)$$

where $\mathbf{y}_d$ and $\mathbf{y}_s$ are the *purely deterministic* (p.d.) and the *purely non-deterministic* (p.n.d.) components, the latter with an absolutely continuous spectrum and a log-integrable spectral density; see e.g. [Rozanov \(1967\)](#). If $\mathbf{y}$ admits finite-dimensional realizations it can be described by a minimal state space model of the form,

$$\begin{bmatrix} \mathbf{x}(t+1) \\ \mathbf{z}(t+1) \end{bmatrix} = \begin{bmatrix} A_d & 0 \\ 0 & A_s \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{z}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ K \end{bmatrix} \mathbf{e}(t) \quad (2.2a)$$

$$\mathbf{y}(t) = \begin{bmatrix} C_d & C_s \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{z}(t) \end{bmatrix} + \mathbf{e}(t) \quad (2.2b)$$

where the undriven subsystem with p.d. output $\mathbf{y}_d(t) := C_d \mathbf{x}(t)$, described by an observable pair (A_d, C_d) has a positive definite initial state covariance matrix $P_d = \mathbb{E} \mathbf{x}(0) \mathbf{x}(0)^\top$. The minimal triplet (C_s, A_s, K) describing the p.n.d. component $\mathbf{y}_s(t) = C_s \mathbf{z}(t) + \mathbf{e}(t)$ originates a stable minimum phase transfer function $I + C_s(zI - A_s)^{-1}K$. Here $\mathbf{e}$ will be taken to be the innovation process of $\mathbf{y}_s$ (and hence of $\mathbf{y}$ as well), having a positive definite covariance matrix $\Delta := \mathbb{E} \mathbf{e}(t) \mathbf{e}(t)^\top$ which we shall write in factorized form as $\Delta = DD^\top$ with a nonsingular factor D . In the following we shall need a.s. convergence of the sample second order moments of the p.n.d. output component. To ensure this (second order ergodicity) we may assume that $\mathbf{e}$ is a stationary martingale difference with finite fourth order moments. See [Hannan and Deistler \(1988\)](#) or [Peternell et al. \(1995, Section 3\)](#).

By stationarity and minimality the two block-vector components of the initial state $[\mathbf{x}(0)^\top \mathbf{z}(0)^\top]^\top$ of (2.2) must be uncorrelated. Each has a positive definite covariance matrix, satisfying the Lyapunov equations

$$P_d = AP_dA^\top, \quad P_s = A_s P_s A_s^\top + KK^\top. \quad (2.3)$$

We shall denote by d the dimension of the p.d. subsystem and by p the dimension of the p.n.d. subsystem in (2.2) and let $n := d + p$. Occasionally we shall use the more compact notations

$$C = \begin{bmatrix} C_d & C_s \end{bmatrix}, \quad P := \text{diag} \{P_d, P_s\}, \quad A := \text{diag} \{A_d, A_s\}. \quad (2.4)$$

A special class of signals (2.1) is obtained when the p.n.d. component is white noise; i.e. $\mathbf{y}(t) = \mathbf{y}_d(t) + \mathbf{e}(t)$. Due to their importance in diverse applications, especially frequency estimation, a huge literature has been devoted to the identification of these signals; see e.g. the book [\(Stoica & Moses, 2005\)](#) and the references therein.

Since $\mathbf{y}_d$ and $\mathbf{y}_s$ are completely uncorrelated, the covariance function of the output process $\mathbf{y}$ splits into its p.d. and p.n.d. components

$$\Lambda(\tau) := \mathbb{E} \mathbf{y}(t + \tau) \mathbf{y}(t)^\top = \Lambda_d(\tau) + \Lambda_s(\tau)$$

with the p.n.d. part having the well-known representation, see e.g. [Anderson \(1969\)](#),

$$\begin{cases} \Lambda_s(\tau) = C_s A_s^{\tau-1} \bar{C}_s^\top & \text{for } \tau = 1, 2, \dots \\ \Lambda_s(0) = C_s P_s C_s^\top + DD^\top & \text{for } \tau = 0 \end{cases} \quad (2.5)$$

where $\bar{C}_s^\top = A_s P_s C_s^\top + KD^\top$. The structure of Λ_d will emerge from the analysis which follows.

We can choose an orthonormal basis in which $P_d = I$, and A_d is an orthogonal (and hence diagonalizable) matrix with complex eigenvalues $e^{\pm i\theta_k}$, $k = 1, \dots, \nu$ and possibly real eigenvalues at $\theta_0 = 0$ and $\theta_{\nu+1} = \pi$. Hence A_d is similar to a block-diagonal real matrix

$$A_d = \text{diag} \left\{ I_{n_0}, \begin{bmatrix} \cos \theta_1 I_{n_1} & -\sin \theta_1 I_{n_1} \\ \sin \theta_1 I_{n_1} & \cos \theta_1 I_{n_1} \end{bmatrix}, \dots, \begin{bmatrix} \cos \theta_\nu I_{n_\nu} & -\sin \theta_\nu I_{n_\nu} \\ \sin \theta_\nu I_{n_\nu} & \cos \theta_\nu I_{n_\nu} \end{bmatrix}, -I_{n_{\nu+1}} \right\} \quad \theta_k \neq \theta_j \quad (2.6)$$

where $n_1, \dots, n_\nu$ are the multiplicities of the complex eigenvalues $e^{\pm i\theta_k}$ appearing in conjugate pairs and n_0 and $n_{\nu+1}$ are the multiplicities of the real eigenvalues $\lambda = 1$ and $\lambda = -1$, some or both of which may possibly be absent. Observability implies that the output dimension m must be an upper bound for the multiplicity of the eigenvalues. Hence for a scalar process n_0 and $n_{\nu+1}$ are ≤ 1 and there are just ν elementary 2×2 oscillatory blocks each corresponding to one of the ν distinct angular frequencies θ_k , $k = 1, \dots, \nu$, which are strictly between $\theta = 0$ and $\theta = \pi$.

The $m \times d$ (where $d = 2 \sum n_k + n_0 + n_{\nu+1}$) matrix C_d splits into blocks $[C_0 \ C_1 \ \dots \ C_\nu \ C_{\nu+1}]$ where C_0 and $C_{\nu+1}$ are $m \times n_0$ and $m \times n_{\nu+1}$ and the C_k , $k = 1, 2, \dots, \nu$ are $m \times 2n_k$. The diagonal block elements in A_d are denoted by A_k . Starting from the complex representation where the matrix A_d is diagonal, each corresponding matrix C_k is (complex and) of full row rank. Thus there exists a collection of rows such that the corresponding submatrix is nonsingular. Using this matrix to transform the basis one can achieve a unity matrix in these rows. Converting back to real matrices then achieves the specific form

$$C_k = \Pi_k \begin{bmatrix} I_{n_k} & 0 \\ H_{k,1} & H_{k,2} \end{bmatrix}, \quad k = 1, \dots, \nu$$

where Π_k is a permutation matrix and the row-block $[H_{k,1} \ H_{k,2}]$ is $(m - n_k) \times 2n_k$. In the scalar case $c_k = [1 \ 0]$ for $k = 1, 2, \dots, \nu$ and 1 otherwise. Returning to the (complex) basis in which A_d is diagonal, it is easy to see that stationarity implies that all the n_k -dimensional complex state subvectors $\mathbf{z}_k(t) := \mathbf{x}_{1,k}(t) + i\bar{\mathbf{x}}_{2,k}(t)$ and $\bar{\mathbf{z}}_k(t) := \mathbf{x}_{1,k}(t) - i\bar{\mathbf{x}}_{2,k}(t)$, $k = 1, 2, \dots, \nu$ and the random vectors $\mathbf{x}_k(t)$, $k = 1, \nu + 1$ must be mutually uncorrelated. This implies in particular that

$$\begin{aligned} \mathbb{E} \{\mathbf{z}_k(0) \bar{\mathbf{z}}_k(0)^\top\} &= \mathbb{E} \{\mathbf{x}_{1,k}(0) \mathbf{x}_{1,k}(0)^\top\} - \mathbb{E} \{\mathbf{x}_{2,k}(0) \mathbf{x}_{2,k}(0)^\top\} \\ &\quad + i (\mathbb{E} \{\mathbf{x}_{1,k}(0) \mathbf{x}_{2,k}(0)^\top\} \\ &\quad + \mathbb{E} \{\mathbf{x}_{2,k}(0) \mathbf{x}_{1,k}(0)^\top\}) = 0 \end{aligned}$$

so that the covariance of $\mathbf{x}_k(0) = [\mathbf{x}_{1,k}(0)^\top \ \mathbf{x}_{2,k}(0)^\top]^\top$ must have the following structure

$$P_k = \mathbb{E} \mathbf{x}_k(0) \mathbf{x}_k(0)^\top = \begin{bmatrix} \Sigma_k & M_k \\ M_k^\top & \Sigma_k \end{bmatrix} \quad M_k = -M_k^\top$$

for $k = 1, 2, \dots, v$. Moreover the overall state covariance matrix $P_d := \mathbb{E}\{\mathbf{x}(t)\mathbf{x}(t)^\top\}$ must be block-diagonal

$$P_d = \text{diag}\{P_0, P_1, \dots, P_v, P_{v+1}\} \quad (2.7)$$

where, for $k = 1, \dots, v$, P_k is $2n_k \times 2n_k$ while P_0 and P_{v+1} are n_0 and n_{v+1} dimensional. Clearly, if $n_k = 1$, $M_k = 0$. Naturally each P_k satisfies a Lyapunov-like equation $P_k = A_k P_k A_k^\top$ and since A_d is orthogonal, P_d and A_d commute (in the chosen basis).

The covariance function of a vector p.d. process can be represented in the form,

$$A_d(\tau) := \mathbb{E}\mathbf{y}_d(t + \tau)\mathbf{y}_d(t)^\top = C_d A_d^{\tau-1} \bar{C}_d^\top, \quad \tau \geq 0, \quad (2.8)$$

where $\bar{C}_d^\top := A_d P_d C_d^\top$ with $P_d := \mathbb{E}\mathbf{x}(0)\mathbf{x}(0)^\top$ positive definite. Note that, since in the canonical basis A_d and P_d commute and P_d is positive definite, the observability of (A_d, C_d) implies the reachability of the pair $(A_d, \bar{C}_d^\top)$. This property is obviously independent of the choice of basis so that the factors in (2.8) define a minimal realization. These structural properties, except perhaps the block-diagonal structure of P_d , are preserved under a change of basis.

Later we shall need a formula for the limits of sample covariances of signals containing p.d. components. Here, hatted symbols will stand for sample estimates. The proof of the following proposition is suppressed for reasons of space limitations.

Proposition 1. Let $\{v(t); t \in \mathbb{Z}\}$ be a sample path of a second order p.n.d. process of zero mean with a rational spectrum, then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N y_d(t + \tau) v(t)^\top = 0 \quad (2.9)$$

with probability one for every τ . Hence the limit sample covariance of the process (2.1) is the sum,

$$\hat{\Lambda}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N y(t + \tau) y(t)^\top = \hat{\Lambda}_d(\tau) + \Lambda_s(\tau) \quad (2.10)$$

where $\hat{\Lambda}_d$ is the limit sample covariance of $\mathbf{y}_d$.

Limits of purely deterministic sample state covariances are described in the next proposition.

Proposition 2. Let $\{x_k(t), x_j(t); t \in \mathbb{Z}\}$ be sample paths of two sub-vectors of the p.d. state process $\mathbf{x}$ expressed in the basis of (2.6), corresponding to frequencies $\theta_k \neq \theta_j$. Then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N x_k(t + \tau) x_j(t)^\top = 0. \quad (2.11)$$

Moreover $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N x_k(t + \tau) x_k(t)^\top = A_k^\tau \hat{P}_k C_k^\top$ where, for $k = 1, 2, \dots, v$,

$$\hat{P}_k = \frac{1}{2} \left(\begin{bmatrix} x_{1,k}(0) & -x_{2,k}(0) \\ x_{2,k}(0) & x_{1,k}(0) \end{bmatrix} \begin{bmatrix} x_{1,k}(0) & -x_{2,k}(0) \\ x_{2,k}(0) & x_{1,k}(0) \end{bmatrix}^\top \right) \quad (2.12)$$

while for $\theta_k = 0$ or π , $\hat{P}_k = x_k(0)x_k(0)^\top$.

A proof of the first statement follows from the arguments of, e.g. Söderström and Stoica (1989, p. 108). The second statement follows by simple calculations and is omitted. Proposition 2 leads to the expression

$$\hat{\Lambda}_d(\tau) = C_d A_d^\tau \hat{P}_d C_d^\top, \quad (2.13)$$

where $\hat{P}_d := \text{diag}\{\hat{P}_0, \hat{P}_1, \dots, \hat{P}_v, \hat{P}_{v+1}\}$. $\hat{P}_d$ may be called the “limit sample state covariance” of the p.d. subsystem. It should

be regarded as a random quantity depending on the (initial conditions; i.e., on the) particular trajectory of the signal.

Lemma 3. The sample state covariance matrix satisfies the Lyapunov equation $\hat{P}_d = A_d \hat{P}_d A_d^\top$. In the special basis of (2.6) $\hat{P}_d$ and A_d commute.

The proof is a simple calculation and is omitted.

It follows from (2.12) that each $2n_k \times 2n_k$ block matrix $\hat{P}_k$ has almost surely rank equal to two so the matrix $\hat{P}_d$ will be nonsingular (with probability one) if and only if A_d has simple eigenvalues.

Comparing (2.13) with the expression (2.8), of the true covariance of a p.d. process, we note that the only difference lies in the fact that the true state covariance P_d is now substituted by the sample covariance $\hat{P}_d$. Note that, even if (A_d, C_d) is observable, the factorization (2.13) will be minimal only if A_d has simple eigenvalues.

3. Asymptotic covariance matching

Assume that we are in an ideal situation of observing an infinitely long sample trajectory $y_1, y_2, \dots, y_t, \dots$ of the output process $\mathbf{y}$ of a true system of the form (2.2). From these data we form a string of limit sample covariances, $\hat{\Lambda}^k := \{\hat{\Lambda}(\tau); \tau = 0, 1, 2, \dots, 2k+1\}$ where the Cesàro limits $\hat{\Lambda}(\tau)$ exist almost surely and are described in Proposition 1. We assume that the integer k is chosen large enough so that $k \geq n$.

Our goal will be to show that notwithstanding the limit p.d. covariance $\hat{\Lambda}_d(\tau)$ is sample dependent, from the asymptotic covariances $\hat{\Lambda}^k$ we can recover the true parameters of the system, $(C_d, A_d, C_s, A_s, K, D)$ modulo a change of basis.

Lemma 4. Assume that A_d has simple eigenvalues. Then the limit sample covariance function $\hat{\Lambda}$ constructed from a sample trajectory of the system (2.2) admits a minimal realization $\hat{\Lambda}(\tau) = C A^{\tau-1} \hat{\bar{C}}^\top$ $\tau = 1, 2, \dots$, with a sample dependent matrix $\hat{\bar{C}}$ given by

$$\hat{\bar{C}}^\top := \text{diag}\{\hat{P}_d P_d^{-1}, I\} \bar{C}^\top := \hat{T} \bar{C}^\top \quad (3.1)$$

where $\hat{P}_d$ denotes the asymptotic sample covariance of the p.d. state component, $\mathbf{x}$, of system (2.2) and the $n \times n$ matrix $\hat{T}$ is nonsingular and commutes with A .

Proof. From formula (2.10),

$$\hat{\Lambda}(\tau) = \hat{\Lambda}_d(\tau) + \Lambda_s(\tau) = C_d A_d^{\tau-1} \hat{P}_d C_d^\top + C_s A_s^{\tau-1} \bar{C}_s^\top,$$

where $\bar{C}_s^\top = A_s P_s C_s^\top + K D^\top$. The expression (3.1) follows from

$$\begin{aligned} \hat{\bar{C}}^\top &= \begin{bmatrix} A_d & 0 \\ 0 & A_s \end{bmatrix} \begin{bmatrix} \hat{P}_d & 0 \\ 0 & P_s \end{bmatrix} \begin{bmatrix} C_d^\top \\ C_s^\top \end{bmatrix} + \begin{bmatrix} 0 \\ K \end{bmatrix} D^\top \\ &= \begin{bmatrix} \hat{P}_d P_d^{-1} & 0 \\ 0 & I \end{bmatrix} \left\{ \begin{bmatrix} A_d & 0 \\ 0 & A_s \end{bmatrix} \begin{bmatrix} P_d & 0 \\ 0 & P_s \end{bmatrix} \begin{bmatrix} C_d^\top \\ C_s^\top \end{bmatrix} + \begin{bmatrix} 0 \\ K \end{bmatrix} D^\top \right\} \\ &= \begin{bmatrix} \hat{P}_d P_d^{-1} & 0 \\ 0 & I \end{bmatrix} \bar{C}^\top \end{aligned}$$

and is valid in the special canonical basis of the p.d. subsystem discussed in Section 2. Since in this basis both P_d^{-1} and $\hat{P}_d$ commute with A_d , we have $A_d \hat{P}_d = \hat{P}_d P_d^{-1} A_d P_d$ and hence $\hat{T}$ commutes with A . $\square$

It follows from this lemma that $A^\tau \hat{C}^\top = A^\tau \hat{T} \bar{C}^\top = \hat{T} A^\tau \bar{C}^\top$ for all $\tau \geq 1$ and there exist a rank n factorization of the asymptotic sample Hankel matrix

$$\hat{\mathbf{H}}(k+1) := \begin{bmatrix} \hat{\Lambda}(1) & \hat{\Lambda}(2) & \cdots & \hat{\Lambda}(k+1) \\ \hat{\Lambda}(2) & \hat{\Lambda}(3) & \cdots & \hat{\Lambda}(k+2) \\ \hat{\Lambda}(3) & & \cdots & \cdots \\ \vdots & & \cdots & \vdots \\ \hat{\Lambda}(k+1) & \hat{\Lambda}(k+2) & \cdots & \hat{\Lambda}(2k+1) \end{bmatrix} \quad (3.2)$$

of the form

$$\hat{\mathbf{H}}(k+1) = \mathbf{\Omega}(k+1) \hat{T} \bar{\mathbf{\Omega}}(k+1)^\top \quad (3.3)$$

where $\mathbf{\Omega}(k+1)$ and $\bar{\mathbf{\Omega}}(k+1)$ are the extended observability and reconstructability matrices of the system generating the data. Hence by chopping off the last blocks of m rows from $\mathbf{\Omega}(k+1)$ and $\bar{\mathbf{\Omega}}(k+1)$ one obtains corresponding factorizations of the $m(k+1) \times mk$, $mk \times mk$ and $mk \times m(k+1)$ asymptotic Hankel sub-matrices denoted as follows:

$$\hat{\mathbf{H}}(k+1, k) = \mathbf{\Omega}(k+1) \hat{\mathbf{\Omega}}^\top(k) \quad (3.4a)$$

$$\hat{\mathbf{H}}(k) = \mathbf{\Omega}(k) \hat{\mathbf{\Omega}}^\top(k) \quad (3.4b)$$

$$\hat{\mathbf{H}}(k, k+1) = \mathbf{\Omega}(k) \hat{\mathbf{\Omega}}^\top(k+1) \quad (3.4c)$$

where the factors $\hat{\mathbf{\Omega}}$ include the sample dependent term $\hat{T}$. If $k \geq n$ all these factorizations are of rank n , Kalman's equal rank condition (Kalman, 1971) is satisfied, and hence $(C, A, \hat{C}^\top)$ is the unique (modulo similarity) partial realization of the finite sequence $\hat{\Lambda}^k$. This leads to the following.

Theorem 5. Assume that $k \geq n$. Then, if and only if A_d has simple eigenvalues, from the asymptotic sample covariances $\hat{\Lambda}^k$ one can uniquely recover the true parameters $(C_d, A_d, C_s, A_s, K, D)$ of the system (2.2) modulo a change of basis.

Proof. In fact, assume that A_d has simple eigenvalues; then by partial realization we can map $\hat{\Lambda}^k$ into a minimal triplet $(C, A, \hat{C}^\top)$ of dimension n with $\hat{C} = \hat{C}(\omega)$ depending on the sample data. Decompose A in block diagonal form $A = A_d \oplus A_s$ as in (2.2) with A_d oscillatory and A_s with eigenvalues strictly inside the unit circle and partition $C = [C_d \ C_s]$ conformably. Now, because of (3.1), in any such basis we have

$$\hat{C}^\top = \begin{bmatrix} \hat{C}_d^\top \\ \hat{C}_s^\top \end{bmatrix} = \begin{bmatrix} A_d \hat{P}_d C_d^\top \\ A_s P_s C_s^\top + K D^\top \end{bmatrix}$$

modulo similarity. Since $\hat{P}_d$ and A_d commute, we have

$$\begin{bmatrix} \hat{C}_d^\top & A_d^\top \hat{C}_d^\top & \cdots & (A_d^{d-1})^\top \hat{C}_d^\top \\ \hat{C}_s^\top & A_s^\top \hat{C}_s^\top & \cdots & (A_s^{d-1})^\top \hat{C}_s^\top \end{bmatrix} = \hat{P}_d A_d \begin{bmatrix} C_d^\top & A_d^\top C_d^\top & \cdots & (A_d^{d-1})^\top C_d^\top \end{bmatrix} \quad (3.5)$$

and we can recover the sample state covariance matrix $\hat{P}_d$ from $(C_d, A_d, \hat{C}_d^\top)$. Next, the true covariance, $\Lambda_s(0)$, of the purely non-deterministic component of the output process can be obtained by subtracting the first summand in $\hat{\Lambda}(0) = C_d \hat{P}_d C_d^\top + \Lambda_s(0)$. Finally from $C_s, A_s, \bar{C}_s, \Lambda_s(0)$, the parameters K and D can be computed by solving an algebraic Riccati equation. Hence we can recover the system parameters uniquely modulo similarity.

Conversely, if A_d has multiple eigenvalues we have $\text{rank} \hat{P}_d < n$ so the factorizations (3.4) have rank strictly smaller than n and a minimal realization of $\hat{\Lambda}^k$ cannot have order equal to n . $\square$

4. Subspace identification as partial realization

In this section, we shall recall the basic steps of a subspace identification algorithm for time series which will be generically referred to as CCA (Canonical Correlation Analysis) algorithm. CCA is actually a first step common to many subspace algorithms to obtain a factorization of the sample Hankel matrix and simultaneously accomplish order estimation. From this factorization some procedures compute estimates of the matrices $(C, A, \bar{C})$ by (approximate) partial realization by solving certain “shift-invariance” equations. This procedure is based on a long series of well-known early contributions (Akaike, 1976; Aoki, 1990; Larimore, 1983; Zeiger & McEwan, 1972). Others (Larimore, 1990; van Overschee & De Moor, 1993) compute instead an estimate of the state using the canonical variables and mimic the steps of stochastic realization (the “state approach”). See Bauer (2005) for a historical account and Lindquist and Picci (1996), Lindquist and Picci (1996) for a detailed explanation of the theoretical background on which this procedure is based. It is a little known fact that both procedures lead to the same estimates. The equivalence of the stochastic realization procedure with partial realization is implicit in Lindquist and Picci (1996, Theorem 7.1), Lindquist and Picci (1996), and explicitly proven in Chiuso and Picci (2004, Section 4.2) although in a more general context where input signals are also present. For the purpose of this paper it will be convenient to refer to the partial realization version of the algorithm.

The first step of the CCA procedure is to regroup the output data, say $y_0, y_1, y_1, \dots, y_t, \dots, y_{N'}$ (after subtracting off the sample mean) into truncated tail matrices with $N+1$ columns

$$Y_N(t) := [y_t, y_{t+1}, \dots, y_{t+N}] \quad t = 0, 1, \dots$$

By Propositions 1 and 2, the sample covariances

$$\hat{\Lambda}_N(i, j) = \mathbb{E}_N \{Y_N(i) Y_N(j)^\top\} := \frac{1}{N+1} Y_N(i) Y_N(j)^\top \quad (4.1)$$

converge for $N \rightarrow \infty$ to the limit

$$\lim_{N \rightarrow \infty} \mathbb{E}_N \{Y_N(i) Y_N(j)^\top\} = \hat{\Lambda}_d(i-j) + \Lambda_s(i-j)$$

for all $i, j \geq 0$. Hence, for a stationary signal containing almost-periodic components the limit on the right is not the expectation but just the (sample dependent) asymptotic covariance $\hat{\Lambda}(i-j)$.

Fixing a present time $t = k$, the tails may be stacked into “past” and “future” data matrices at time k , of dimension $m(k+1) \times (N+1)$,

$$Y_{[0,k]} = \begin{bmatrix} Y_N(k) \\ \vdots \\ Y_N(1) \\ Y_N(0) \end{bmatrix}, \quad Y_{[k+1,2k+1]} = \begin{bmatrix} Y_N(k+1) \\ Y_N(k+2) \\ \vdots \\ Y_N(2k+1) \end{bmatrix}. \quad (4.2)$$

In the identification literature the lengths of the past and future horizons are often taken to be different “design parameters”. Although this could easily be accommodated, there is no real loss of generality in keeping horizons of equal length $k+1$ as defined above.

From the past and future data matrices $Y_{[0,k]}, Y_{[k+1,2k+1]}$, one forms the sample Hankel matrix

$$\hat{\mathbf{H}}_N(k+1) := \mathbb{E}_N \{Y_{[k+1,2k+1]} Y_{[0,k]}^\top\} \quad (4.3)$$

of which we shall consider the submatrices

$$\hat{\mathbf{H}}_N(k+1, k) := \mathbb{E}_N \{Y_{[k+1,2k+1]} Y_{[1,k]}^\top\}, \quad (4.4a)$$

$$\hat{\mathbf{H}}_N(k) := \mathbb{E}_N \{Y_{[k+1,2k]} Y_{[1,k]}^\top\}, \quad (4.4b)$$

$$\hat{\mathbf{H}}_N(k, k+1) := \mathbb{E}_N \{Y_{[k+1,2k]} Y_{[0,k]}^\top\}, \quad (4.4c)$$

of dimension $m(k+1) \times mk$, $mk \times mk$ and $mk \times m(k+1)$, which, for $N \rightarrow \infty$ tend to matrices with a Hankel structure. Even if for N finite do not have a Hankel structure, they are usually called *sample Hankel matrices* anyway.

As a first step, $\hat{\mathbf{H}}_N(k+1)$ and (4.4a) are *normalized* by computing the Cholesky factors, L_{k+1}^- and L_{k+1}^+ , of the sample covariances $T_{k+1}^- := \mathbb{E}_N\{Y_{[0,k]}Y_{[0,k]}^\top\}$ and $T_{k+1}^+ := \mathbb{E}_N\{Y_{[k+1,2k+1]}Y_{[k+1,2k+1]}^\top\}$, and defining $\hat{H}_N(k+1) := (L_{k+1}^+)^{-1}\hat{\mathbf{H}}_N(k+1)(L_{k+1}^-)^{-\top}$, $\hat{H}_N(k+1, k) := (L_{k+1}^+)^{-1}\hat{\mathbf{H}}_N(k+1, k)(L_k^-)^{-\top}$ and $\hat{H}_N(k, k+1) := (L_k^+)^{-1}\hat{\mathbf{H}}_N(k, k+1)(L_{k+1}^-)^{-\top}$ where L_k^- and L_k^+ are submatrices of L_{k+1}^- and L_{k+1}^+ defined coherently with the ordering in (4.2).

Next, factorizations of these matrices are obtained by a truncated SVD technique on the normalized sample Hankel matrix $\hat{H}_N(k+1)$,

$$\hat{H}_N(k+1) = [\hat{U}_N \quad \tilde{U}_N] \begin{bmatrix} \hat{\Sigma}_N & 0 \\ 0 & \tilde{\Sigma}_N \end{bmatrix} \begin{bmatrix} \hat{V}_N^\top \\ \tilde{V}_N^\top \end{bmatrix} \quad (4.5)$$

by keeping the first $\hat{n}$ “significant” singular values in $\hat{\Sigma}_N$ while discarding $\tilde{\Sigma}_N$ so as to form the best approximant of rank $\hat{n}$,

$$\hat{\mathbb{H}}_N(k+1) := \hat{U}_N \hat{\Sigma}_N \hat{V}_N^\top = \hat{\Omega}_N(k+1) \hat{\Omega}_N(k+1)^\top \quad (4.6)$$

where $\hat{\Omega}_N(k+1) := \hat{U}_N \hat{\Sigma}_N^{1/2}$ and $\hat{\Omega}_N(k+1) := \hat{V}_N \hat{\Sigma}_N^{1/2}$ are $m(k+1) \times \hat{n}$ both of full column rank.

Let $\hat{\Omega}_N(k)$ and $\hat{\Omega}_N^\top(k)$ be the factors $\hat{\Omega}_N(k+1)$, $\hat{\Omega}_N(k+1)^\top$ in (4.6) with the last block of m rows deleted. If k is large enough both $\hat{\Omega}_N(k)$ and $\hat{\Omega}_N^\top(k)$ will generically still have rank $\hat{n}$. In this case, all submatrices in the factorizations

$$\hat{\mathbb{H}}_N(k) := \hat{\Omega}_N(k) \hat{\Omega}_N^\top(k), \quad (4.7a)$$

$$\hat{\mathbb{H}}_N(k+1, k) := \hat{\Omega}_N(k+1) \hat{\Omega}_N^\top(k), \quad (4.7b)$$

$$\hat{\mathbb{H}}_N(k, k+1) := \hat{\Omega}_N(k) \hat{\Omega}_N^\top(k+1), \quad (4.7c)$$

have rank $\hat{n}$.

Next, form the unnormalized factors

$$\hat{\Omega}_N(k) := L_k^+ \hat{\Omega}_N(k) \quad \hat{\Omega}_N^\top(k) := L_k^- \hat{\Omega}_N^\top(k), \quad (4.8a)$$

$$\hat{\Omega}_N(k+1) := L_{k+1}^+ \hat{\Omega}_N(k+1), \quad (4.8b)$$

$$\hat{\Omega}_N^\top(k+1) := L_{k+1}^- \hat{\Omega}_N^\top(k+1) \quad (4.8c)$$

and solve the shift-invariance equations

$$\hat{\Omega}_N(k+1) = \begin{bmatrix} C \\ \hat{\Omega}_N(k)A \end{bmatrix} \quad (4.9a)$$

$$\hat{\Omega}_N^\top(k+1) = \begin{bmatrix} \tilde{C} \\ \hat{\Omega}_N^\top(k)\bar{A}^\top \end{bmatrix} \quad (4.9b)$$

to get estimates, denoted by $\hat{C}_N$, $\hat{A}_N$, $\hat{\bar{C}}_N$, of $(C, A, \bar{C})$. The estimates $\hat{C}_N$ and $\hat{\bar{C}}_N$ are simply the first block-rows of $\hat{\Omega}_N(k+1)$ and $\hat{\Omega}_N^\top(k+1)$ while an estimate of A can be obtained by solving in the least-squares sense the overdetermined system

$$\downarrow \hat{\Omega}_N(k+1) = \hat{\Omega}_N(k)A \quad (4.10)$$

where $\downarrow :=$ multiplication by $[0_{mk \times m} \quad I_{mk}]$, is the “chop off” operator which acts on matrices by deleting the first block of m rows.

When rank $\hat{\Omega}_N(k) = \hat{n}$, we shall convene to choose

$$\hat{A}_N := [\hat{\Omega}_N(k)^\top \hat{\Omega}_N(k)]^{-1} \hat{\Omega}_N(k)^\top (\downarrow \hat{\Omega}_N(k+1)) \quad (4.11)$$

otherwise the least square estimate can be defined in terms of the Moore–Penrose pseudoinverse. A similar formula can be given for the estimate of $\bar{A}$.

5. Proof of consistency

Since the covariance function Λ can be parametrized by a whole family of minimal matrix triplets $(C, A, \bar{C}^\top)$ mutually equivalent modulo similarity, a sequence of minimal estimates $(\hat{C}_N, \hat{A}_N, \hat{\bar{C}}_N)$ is called *consistent* if there is a sequence of nonsingular matrices $\{T_N\}$ such that $(\hat{C}_N T_N, T_N^{-1} \hat{A}_N T_N, T_N^{-1} \hat{\bar{C}}_N^\top)$ converges for $N \rightarrow \infty$ to a triplet realizing the true covariance of the system which generates the data. We shall call this *convergence modulo similarity* for short. An equivalent definition can be given in terms of canonical forms representing the equivalence classes modulo similarity.

Unfortunately even this relaxed notion of consistency does in general not apply to the parametric structure of the covariance function of a p.d. model. Any covariance function of this kind depends on three matrix parameters, say $(C_d, A_d, \bar{C}_d)$, or equivalently (C_d, A_d, P_d) ; see (2.8) where either $\bar{C}_d$ or the initial state covariance matrix P_d are parameters which cannot be consistently estimated from the data. In particular P_d is an ensemble average of different state sample amplitudes which we cannot have any information about from a single sample path, even if infinite data length was available. For this reason we shall use a weaker notion of consistency for p.d. models, identifying a p.d. model with its state-output equations, parametrized by an observable pair (C_d, A_d) . We shall henceforth just say that a sequence of p.d. system estimates $(\hat{C}_{d,N}, \hat{A}_{d,N})$ is *consistent* if it converges modulo similarity to (C_d, A_d) . This can be equivalently stated by saying that the limit of the pair $(\hat{C}_{d,N}, \hat{A}_{d,N})$ expressed in a suitable observability canonical form, is equal to the “true” observable pair (C_d, A_d) expressed in the same canonical form. This is for example enough to guarantee a.s. convergence of frequency estimates.

Theorem 6. Assume that the p.d. subsystem of (2.2) has simple eigenvalues and that k is chosen large enough so that $k \geq n$. Assume also that $\hat{n}$ is a consistent estimate of the true order n of the system. Then the finite sample estimates $(\hat{C}_N, \hat{A}_N, \hat{\bar{C}}_N)$ defined in Section 4, converge modulo similarity and with probability one to a minimal realization, $(C, A, \bar{C}^\top)$, of the limit sample covariance $\hat{\Lambda}$ of the system.

Proof. The idea of the proof is to show that the approximate rank factorizations (4.7) converge to the exact rank factorizations of the asymptotic normalized Hankel matrix $\hat{H}(k+1)$ constructed as in (3.2), with the limit covariance sequence $\hat{\Lambda}^k$ of Section 3. Since $\hat{\Lambda}_N^k \rightarrow \hat{\Lambda}^k$, consistency will follow by a continuity argument and by Theorem 5.

Now, because of a.s. consistency of the order estimator there is a (random) sample size N_0 such that $\hat{n}$ will be equal to the true orders n of the system (2.2), indefinitely for all $N > N_0$. In the following we shall assume that $N > N_0$. Then the sample truncated singular value matrix $\hat{\Sigma}_N$ computed by the algorithm will be $n \times n$, and the corresponding (normalized) observability and constructability factors, $\hat{\Omega}_N(k+1)$, $\hat{\Omega}_N^\top(k+1)$, of $\hat{H}_N(k+1)$ will have n columns, so that all involved factorizations will be of rank n .

Lemma 7. Under the stated assumptions, there is a sequence of nonsingular matrices $\{T_N\}$ such that

$$\hat{\mathbf{Z}}_N(k+1)T_N \rightarrow \hat{\mathbf{Z}}(k+1), \quad (5.1)$$

$$\hat{\mathbf{Z}}_N(k+1)T_N^{-\top} \rightarrow \hat{\mathbf{Z}}(k+1),$$

$$\hat{\mathbf{Z}}_N(k)T_N \rightarrow \hat{\mathbf{Z}}(k), \quad \hat{\mathbf{Z}}_N(k)T_N^{-\top} \rightarrow \hat{\mathbf{Z}}(k), \quad (5.2)$$

almost surely, where the $\mathbf{Z}$'s without subscripts are rank n factors of the asymptotic unnormalized sample Hankel matrices in (3.4).

Proof. By consistency of the order estimator, for all $N > N_0$, the truncated SVD approximation, $\hat{\mathbf{H}}_N(k+1)$, has rank n equal to the rank of the asymptotic Hankel matrix $\hat{H}(k+1)$. Then by a well-known estimate (see e.g. Golub and van Loan (1983)),

$$\|\hat{\mathbf{H}}_N(k+1) - \hat{H}(k+1)\| \leq \hat{\sigma}_N(n+1)$$

where the $\hat{\sigma}_N(j)$ are the singular values of $\hat{H}_N(k+1)$ (of course depending on N). Now, since the singular values are a continuous function of the relative matrix (see e.g. Stewart and Sun (1990, p. 204)), the singular values of index j greater than n of $\hat{H}(k+1)$ are all zero, and $\lim_{N \rightarrow \infty} \hat{H}_N(k+1) = \hat{H}(k+1)$, it follows that all singular values of $\hat{H}_N(k+1)$ of index greater than n must tend to zero for $N \rightarrow \infty$ (although $\hat{H}_N(k+1)$ may have rank $> n$ for all N) and therefore $\hat{\mathbf{H}}_N(k+1) \rightarrow \hat{H}(k+1)$ for $N \rightarrow \infty$.

Consider the SVD factorization $\hat{H}(k+1) = \hat{U} \hat{\Sigma} \hat{V}^\top$ where $\hat{\Sigma}$ is $n \times n$ and let $\hat{U} := \text{span}\{\hat{U}\}$ and $\hat{V} := \text{span}\{\hat{V}\}$ denote the column spaces of $\hat{U}$ and $\hat{V}$. Since $\hat{\mathbf{H}}_N(k+1)$ converges to $\hat{H}(k+1)$, the eigenspaces $\hat{\mathbf{U}}_N := \text{span}\{\hat{U}_N\}$ and $\hat{\mathbf{V}}_N := \text{span}\{\hat{V}_N\}$ must converge in the gap metric to $\hat{U}$ and $\hat{V}$, respectively, see Stewart and Sun (1990, p. 260)², the gap metric on subspaces of a Hilbert space being defined as

$$\gamma(\mathbf{X}, \mathbf{Y}) := \|\Pi_{\mathbf{X}} - \Pi_{\mathbf{Y}}\|$$

where $\Pi_{\mathbf{X}}$ denotes an orthogonal projection onto $\mathbf{X}$ and $\|\cdot\|$ is the induced operator norm. Convergence in the gap metric of $\hat{\mathbf{U}}_N$ to $\hat{U}$ then implies that $\Pi_{\hat{\mathbf{U}}_N} \hat{U}$ must converge to $\Pi_{\hat{U}} \hat{U} = \hat{U}$. Since, in the orthonormal basis given by the columns of $\hat{U}_N$, we have $\Pi_{\hat{\mathbf{U}}_N} \hat{U} = \hat{U}_N \hat{T}_N$, where $\hat{T}_N := \hat{U}_N^\top \hat{U}$, we have $\hat{U}_N \hat{T}_N \rightarrow \hat{U}$, where $\{\hat{T}_N\}$ is a sequence of $n \times n$ matrices which are nonsingular for N large enough. Non-singularity for N large follows from the convergence $\hat{\mathbf{U}}_N \rightarrow \hat{U}$ in the gap metric and can also be checked directly as any nonzero vector a in the nullspace of $\hat{U}_N^\top \hat{U}$ would correspond to a vector $\hat{U}a \in \hat{U}$ orthogonal to $\hat{\mathbf{U}}_N$. Let now $T_N := \hat{\Sigma}_N^{-1/2} \hat{T}_N \hat{\Sigma}_N^{1/2}$ so that $\hat{\mathbf{Z}}_N(k+1)T_N = \hat{U}_N \hat{T}_N \hat{\Sigma}_N^{1/2}$ converges to $\hat{U} \hat{\Sigma}^{1/2} = \hat{\mathbf{Z}}(k+1)$.

Now, since $\hat{\mathbf{H}}_N(k+1) = \hat{\mathbf{Z}}_N(k+1)T_N T_N^{-1} \hat{\mathbf{Z}}_N(k+1)^\top$ converges to $\hat{H}(k+1) = \hat{\mathbf{Z}}(k+1)\hat{\Sigma}(k+1)^\top$ and $\hat{\mathbf{Z}}_N(k+1)T_N \rightarrow \hat{\mathbf{Z}}(k+1)$, all factors having (for N large enough) linearly independent columns, it follows that $T_N^{-1} \hat{\mathbf{Z}}_N(k+1)^\top$ converges to $\hat{\Sigma}(k+1)^\top$.

We may equivalently say that the convergence $\hat{\mathbf{Z}}_N(k+1) \rightarrow \hat{\mathbf{Z}}(k+1)$ and $\hat{\mathbf{Z}}_N(k+1) \rightarrow \hat{\mathbf{Z}}(k+1)$ takes place in the gap metric. Moreover, since $\hat{\mathbf{H}}_N(k) \rightarrow \hat{H}(k)$, $\hat{\mathbf{H}}_N(k+1, k) \rightarrow \hat{H}(k+1, k)$, $\hat{\mathbf{H}}_N(k, k+1) \rightarrow \hat{H}(k, k+1)$ almost surely for $N \rightarrow \infty$, the factors, $\hat{\mathbf{Z}}_N(k)$ and $\hat{\mathbf{Z}}_N(k)$, in (4.7), tend in the gap metric almost surely to upper truncations of dimension $mk \times n$, of the left and right factors, $\hat{\mathbf{Z}}(k+1)$ and $\hat{\mathbf{Z}}(k+1)$, of $\hat{H}(k+1)$.

Consequently, since the Cholesky factors $L_k^\pm, L_{k+1}^\pm$ converge almost surely as $N \rightarrow \infty$, the unnormalized factors (4.8) also converge in the gap metric when $N \rightarrow \infty$, so there is a sequence of nonsingular matrices $\{T_N\}$ such that (5.1) and (5.2) are satisfied almost surely. $\square$

It is now easy to check that the solution of the shift invariance equations (4.9) in which $\hat{\mathbf{Z}}_N(k+1)$ and $\hat{\mathbf{Z}}_N(k)$ are substituted by $\hat{\mathbf{Z}}_N(k+1)T_N, \hat{\mathbf{Z}}_N(k)T_N$ etc. is just the similarity transform $(\hat{C}_N T_N, T_N^{-1} \hat{A}_N T_N, T_N^{-1} \hat{C}_N^\top)$ of the original estimates, provided N is large enough so that $\text{rank } \hat{\mathbf{Z}}_N(k) = n$. The above immediately implies that $\hat{C}_N T_N \rightarrow C, T_N^{-1} \hat{A}_N T_N \rightarrow A$ and $T_N^{-1} \hat{C}_N^\top \rightarrow \hat{C}^\top$. This is of course due to the fact that the mapping from $(\hat{\mathbf{Z}}_N T_N, T_N^{-1} \hat{\mathbf{Z}}_N^\top)$ to $(\hat{C}_N T_N, T_N^{-1} \hat{A}_N T_N, T_N^{-1} \hat{C}_N^\top)$ is continuous for all matrices where $\hat{\mathbf{Z}}_N(k)T_N$ has full rank n . $\square$

Note that this proof works for a (large enough but) fixed future/past regression horizon k and does not require that k tends to infinity at a certain rate with the sample length N , as required in a large part of the current literature on the asymptotic behavior of subspace methods. This seems to fit the spirit of the discussion about the need of the condition $i \rightarrow \infty$ in van Overschee and De Moor (1993, Section 9). So far we have shown convergence of the overall covariance estimation procedure. To complete the picture we should discuss how to extract consistent (mod similarity) estimates of the (C_d, A_d) and $(C_s, A_s, \bar{C}_s)$ parameters. In the following we shall take A in block diagonal form $A = \text{diag}\{A_d, A_s\}$ as in (2.2). By Theorem 6 there is a sequence of invertible matrices T_N such that $T_N^{-1} \hat{A}_N T_N - A \rightarrow 0$ almost surely as $N \rightarrow \infty$. Here we shall need to assume that this happens at a certain rate $O(f(N))$ with $f(N) \rightarrow 0$ as $N \rightarrow \infty$. Typically when A has only eigenvalues of modulus strictly less than one, $f(N)$ will tend to zero as $1/\sqrt{N}$, see e.g. Bauer (2005); Chiuso and Picci (2004) and for concreteness we shall assume f to be a function of this class in what follows. The final result will however be independent of the particular form of f . Then, for $N \rightarrow \infty$

$$T_N^{-1} \hat{A}_N T_N = A + \delta A_N / \sqrt{N}$$

with δA_N almost surely bounded. Let $\alpha(N)$ be an integer-valued function which grows more slowly than $\sqrt{N}$, for $N \rightarrow \infty$. Then, from Newton binomial expansion, the second member in the expression

$$\hat{A}_N^{\alpha(N)} = T_N (A + \delta A_N / \sqrt{N})^{\alpha(N)} T_N^{-1}$$

is of the same order of $T_N A^{\alpha(N)} T_N^{-1}$ so that, for $N \rightarrow \infty$, $T_N^{-1} \hat{A}_N^{\alpha(N)} T_N \rightarrow \text{diag}\{A_d^{\alpha(N)}, 0\}$, since all eigenvalues of A_s are in modulus strictly less than one. Hence for $N \rightarrow \infty$ the range space of $\hat{A}_N^{\alpha(N)}$ (which may be computed say by an ordered real Schur decomposition) becomes arbitrarily close in the gap metric to the d -dimensional eigenspace spanned by the eigenvalues of modulus one of A . This eigenspace is a simple invariant subspace of A in the sense of Stewart and Sun (1990, p. 221). Then, by continuity of the eigenvalues and eigenspaces with respect to the relative matrix, see e.g. Stewart and Sun (1990, pp. 236–241), consistent estimates of C_d, A_d (and C_s, A_s) can be found by restricting $\hat{C}_N, \hat{A}_N$ to this invariant subspace or, equivalently, by enforcing shift invariance to the two submatrices $[\hat{\mathbf{Z}}_{d,N}(k+1) \quad \hat{\mathbf{Z}}_{s,N}(k+1)]$ obtained from the corresponding change of basis on $\hat{\mathbf{Z}}_N(k+1)$. The eigenvalues of the estimate $\hat{A}_{d,N}$ in general will not have modulus one. Therefore it is reasonable to add a further step where $\hat{A}_{d,N}$ is replaced by an “equivalent” orthogonal matrix. In the new basis one can

² A theorem of Chatelin quoted in Bauer and Jansson (2000, Lemma 7) can also be used for the same purpose.

impose that the solution of the shift invariance equation for the deterministic A matrix,

$$\downarrow \hat{\Omega}_{d,N}(k+1) = \hat{\Omega}_{d,N}(k)\hat{A}_d$$

should be an orthogonal matrix. This is a *Matrix Procrustes Problem* which can be solved by SVD (Golub & van Loan, 1983). The estimate $\hat{P}_d$ may then be computed as in (3.5). Note that knowledge of the exact order of convergence $O(1/\sqrt{N})$ is not important as the argument above works for any function $f(N)$ such that $\alpha(N)f(N)$ tends to zero as $N \rightarrow \infty$.

Estimates $\hat{K}_N$ and $\hat{D}_N$ can be computed for finite N by solving a *sample algebraic Riccati equation* depending on the estimated parameters $(\hat{C}_{s,N}, \hat{A}_{s,N}, \hat{\hat{C}}_{s,N}, \hat{\Lambda}_{s,N}(0))$ of the p.n.d. subsystem.

$$X = \hat{A}_{s,N}X\hat{A}_{s,N}^\top + (\hat{\hat{C}}_{s,N}^\top - \hat{A}_{s,N}X\hat{C}_{s,N}^\top) \\ \times (\hat{\Lambda}_{s,N}(0) - \hat{C}_{s,N}X\hat{C}_{s,N}^\top)^{-1}(\hat{\hat{C}}_{s,N}^\top - \hat{A}_{s,N}X\hat{C}_{s,N}^\top)^\top$$

where $\hat{\Lambda}_{s,N}(0) := \hat{\Lambda}_N(0) - \hat{C}_{d,N}\hat{P}_{d,N}\hat{\hat{C}}_{d,N}^\top$. This will converge for $N \rightarrow \infty$ to the true ARE for the p.n.d. subsystem of (2.2). Note however that even if the sequence of estimates $(\hat{C}_{s,N}, \hat{A}_{s,N}, \hat{\hat{C}}_{s,N}, \hat{\Lambda}_{s,N}(0))$ is consistent, there is no guarantee that the estimated system will be positive real (i.e. define a positive semidefinite spectral density) for finite N . In other words, even if the spectrum estimate $\hat{\Phi}_{s,N}(z)$, of the p.n.d. component for $N \rightarrow \infty$ will converges to a positive function, there is no guarantee that for any finite N this estimate should be positive since the sequence of functions on the unit circle $\{\hat{\Phi}_{s,N}(e^{j\theta}); N = 1, 2, \dots\}$ may converge to a nonnegative limit without being everywhere nonnegative as functions of θ , for any N . Therefore the algebraic Riccati equation based on finite sample estimates $(\hat{C}_{s,N}, \hat{A}_{s,N}, \hat{\hat{C}}_{s,N})$ may fail to have a solution for all N . To avoid this nuisance an additional assumption of *strict positive realness (or coercivity) of the true spectrum* is needed. Then the limit when $N \rightarrow \infty$ of $\hat{\Lambda}_{s,N}$ will be a positive (and sample independent) sequence for N large enough. The following result is taken from Lindquist and Picci (1996).

Theorem 8. Assume that the spectral density function, $\Phi_s(z)$, of the p.n.d. component of the process $\mathbf{y}$ is strictly positive definite on the unit circle. Then, under the same assumptions of Theorem 6, the CCA partial realization algorithm applied to a string of sample covariance lags generated by the true system (2.2) will yield consistent estimates of the parameter K and D modulo multiplication from the right by an arbitrary orthogonal matrix.

6. Conclusions

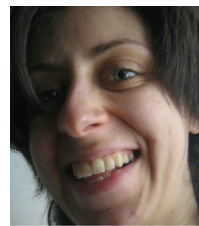
In this paper, we have presented a general proof of strong consistency of subspace identification applied to signals with oscillatory components. Even if these signals are not ergodic and hence the standard consistency arguments based on second-order ergodicity do not apply, their sample covariance converges almost surely (to a sample dependent limit) and this fact can be exploited to show convergence of the estimates of the identifiable system parameters.

Acknowledgments

We are grateful to the associate editor and the referees for their scrupulous reading and for pointing out several errors in earlier versions of this manuscript.

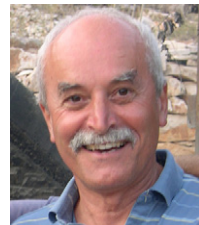
References

- Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In *System identification: advances and case studies* (pp. 27–96). Academic Press.
- Anderson, B. D. O. (1969). The inverse problem of stationary covariance generation. *Journal of Statistical Physics*, 1, 133–147.
- Aoki, M. (1990). *State space modelling of time-series* (2nd ed.). NY: Springer.
- Bauer, D. (2005). Asymptotic properties of subspace estimators. *Automatica*, 41, 359–376.
- Bauer, D., & Jansson, M. (2000). Analysis of the asymptotic properties of the moesp type of subspace algorithms. *Automatica*, 36, 497–509.
- Bissacco, A., Chiuso, A., & Soatto, S. (2007). Classification and recognition of dynamical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 1958–1972.
- Chiuso, A., & Picci, G. (2004). On the ill-conditioning of subspace identification with inputs. *Automatica*, 40(4), 575–589.
- Golub, G., & van Loan, C. (1983). *Matrix computations*. NY: Johns Hopkins UP.
- Hannan, E. J., & Deistler, M. (1988). *The statistical theory of linear systems*. NY: John Wiley.
- Kalman, R. E. (1971). On minimal partial realizations of a linear input/output map. In R. E. Kalman, & N. de Claris (Eds.), *Aspects of network and system theory* (pp. 385–408). Holt Reinhart and Winston.
- Larimore, W. E. (1983). System identification, reduced-order filtering and modeling via canonical variate analysis. *Proceedings of the American Control Conference*, 445–451.
- Larimore, W. E. Canonical variate analysis in identification, filtering and adaptive control. In proc. of 29th decision and control conference (CDC). (pp. 596–604). Honolulu, Hawaii. 1990.
- Lindquist, A., & Picci, G. (1996). Canonical correlation analysis, approximate covariance extension and identification of stationary time series. *Automatica*, 32, 709–733.
- Lindquist, A., & Picci, G. (1996). Geometric methods for state-space identification. In S. Bittanti, & G. Picci (Eds.), *NATO-ASI Series, Identification, adaptation, learning* (pp. 1–69). New York: Springer Verlag.
- Peternell, K. 1995. Identification of linear dynamic systems by subspace and realization-based algorithms. Ph.D. Thesis. Technical University. Vienna.
- Peternell, K., Deistler, M., & Scherrer, W. (1995). Consistency and relative efficiency of of subspace methods. *Automatica*, 31, 1865–1875.
- Peternell, K., Scherrer, W., & Deistler, M. (1996). Statistical analysis of novel subspace identification methods. *Signal Processing*, 52, 161–178.
- Rozanov, Y. A. (1967). *Stationary random processes*. San Francisco: Holden Days.
- Söderström, T., & Stoica, P. (1989). *System identification*. New York: Prentice Hall.
- Stewart, G. W., & Sun, J. (1990). *Matrix perturbation theory*. Boston: Academic Press.
- Stoica, P., & Moses, R. L. (2005). *Spectral analysis of signals*. New Jersey: Prentice-Hall.
- van Overschee, P., & De Moor, B. (1993). Subspace algorithms for the stochastic identification problem. *Automatica*, 29, 649–660.
- Zeiger, H. P., & McEwan, J. A. (1972). Approximate linear realizations of given dimension via Ho's algorithm. *IEEE Transactions on Automatic Control*, AC-19, 153.



Martina Favaro received a Laurea degree in Computer Engineering from the University of Padova, Italy, in 2008. After graduation she had a short time employment as a research assistant with the Department of Information Engineering, University of Padova and was granted an Erasmus scholarship for participation to the International doctoral school: Measuring, modelling and simulation of nonlinear dynamic systems, Department ELEC, Free University of Brussels, Belgium.

She is currently a Ph.D. student with the Department of Computer Science Engineering and Control Systems, University of Brescia, Italy. Her research interests are in the area of system identification and signal processing.



Giorgio Picci is full professor with the University of Padova, Italy, Department of Information Engineering, since 1980. He graduated (cum laude) from the University of Padova in 1967 and since then has held several long-term visiting appointments with various American, Japanese and European universities among which Brown University, M.I.T., the University of Kentucky, Arizona State University, the Center for Mathematics and Computer Sciences (C.W.I.) in Amsterdam, the Royal Institute of Technology, Stockholm Sweden, Kyoto University and Washington University in St. Louis, Mo. He has been contributing to Systems and Control theory mostly in the area of modeling, estimation and identification of stochastic systems and published over 100 papers and edited three books in this area. He has been involved in various joint research projects with industry and state agencies. He has been chairman of the IFAC Technical Committee on Stochastic Systems, past member of the EUCA council, project manager of the Italian team for the Commission of the European Communities Network of Excellence System Identification (ERNSI) and general coordinator of the Commission of European Communities IST project RECSYS, in the fifth Framework Program. Giorgio Picci is a Life Fellow of the IEEE, an IFAC Fellow and a foreign member of the Swedish Royal Academy of Engineering Sciences.