Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Nonstationary consistency of covariance based subspace methods

Albert Benveniste and Laurent Mevel

***Abstract*— In this paper we study nonstationary consistency of subspace methods for eigenstructure identification, *i.e.,* the ability of subspace algorithms to converge to the true eigenstructure despite nonstationarities.**

I. INTRODUCTION

In 1985, Benveniste and Fuchs [6] proved that the Instrumental Variable method and what was called the Balanced Realization method for linear system eigenstructure identification are consistent for certain classes of nonstationary systems. Since this paper, the family of subspace algorithms has been invented [14], [19], [22], [23], [24] and has expanded rapidly. Therefore, we felt it was timely revisiting the results of [6] and generalizing them to subspace methods. To this end, paper [6] had first to be restructured to show up an important intermediate result, which had not been noticed explicitly. Having done this, we show how this key result can be applied in various situations.

There are a number of convergence studies on subspace methods in a stationary context in the literature, see [11], [3], [4], [5], [9], [10] to mention just a few of them. These papers provide deep and technically difficult results including convergence rates. They typically address the problem of identifying all system matrices or the transfer matrix, *i.e.,* both the pole and zero parts of the system. In contrast, the *nonstationary* consistency property that we study here holds for the eigenstructure (the pole part) only and does not apply to the zero part. It is definitely different from the problem considered in [21].

The paper is organized as follows. The problem of nonstationary consistency is stated in Section II. Section III collects the key results of our approach. In Section III-A we show that nonstationary consistency follows if we can construct matrix estimators having the almost factorized form $R_i(N) = CA^iG(N) + o(1)$, where A and C are the state transition matrix and observation matrix for identification, and $o(1) \to 0$ when the sample size N goes to infinity. Historically, subspace algorithms originated from Instrumental Variable methods; we show in Section III-C that it is in turn useful to associate instruments to subspace algorithms. Since limit theorems exist for martingales regardless of stationarity, our approach to dealing with nonstationarity is based on martingale arguments; this is treated in Section III-D. These results allow us to reformulate the design of subspace algorithms ensuring nonstationary consistency for eigenstructure estimation as the quest for suitable instruments, for linear systems; this is explained in Section III-E. Finally, in Section IV, by using the so developed toolbox of theorems and lemmas, we prove nonstationary consistency of some representative subspace algorithms.

Due to space limitations, we focus in this paper on so-called covariance based subspace algorithms. Projection based ones are not covered. The interested reader is referred to the full paper [1] for the treatment of both families of algorithms.

IRISA/INRIA, Campus de Beaulieu, 35042 Rennes cedex, France. email: firstname.lastname@inria.fr

This work was supported by the FliTE Eureka project number 2419.

II. PROBLEM SETTING

Consider the following linear system

$$\left\{ \begin{array}{rcccccc} x_k & = & Ax_{k-1} & + & Bu_k & + & v_k \\ y_k & = & Cx_{k-1} & + & Du_k & + & w_k \end{array} \right. \tag{1}$$

where $k \in \mathbb{Z}$, x is the $\mathbb{R}^n$-valued state, u is the $\mathbb{R}^m$-valued observed input, v and w are unobserved input disturbances, and y is the $\mathbb{R}^q$-valued observed output.

The key point of this work is that the unobserved input disturbances can be *nonstationary*. For instance, they can be white noises having unknown time-varying covariance matrices. For this case, we should rather reformulate system (1) in the following form, which enlightens that y_k itself is nonstationary in a nontrivial way:

$$\left\{ \begin{array}{rcccccc} x_k & = & Ax_{k-1} & + & Bu_k & + & K(k)\nu_k \\ y_k & = & Cx_{k-1} & + & Du_k & + & L(k)\nu_k \end{array} \right. \tag{2}$$

where $\left[\begin{array}{c} K(k) \\ L(k) \end{array} \right] \left[\begin{array}{cc} K^T(k) & L^T(k) \end{array} \right]$ is the time-varying covariance matrix of the excitation noise in (1), and ν_k is a stationary standard white noise. Note that the zero part of the transfer matrix $[u_k\, \nu_k] \mapsto y_k$ is time-varying in this case, so that consistency makes sense only w.r.t. the pole part.

The problem we consider is the *identification of the pair* (C, A) *up to a change of basis in the state space of system* (2). Equivalently, we identify the pairs $(\lambda, C\varphi_\lambda)$, where λ ranges over the set of eigenvalues of A (the poles of system (2)) and φ_λ are a corresponding set of eigenvectors. In words, we consider the problem of *eigenstructure identification.*

Our objective is to show that subspace methods provide consistent estimators of the eigenstructure, also for nonstationary cases as above. Of course, none of the matrices $A, B, C, D, K(k)$, and $L(k)$, are known. Matrices $B, D, K(k)$, and $L(k)$, are regarded as nuisance and are not for identification in this paper.

III. BASIC THEOREMS FOR NONSTATIONARY CONSISTENCY

Throughout this paper, for t_N a nondecreasing sequence of positive real nombers, $o(t_N)$ generically denotes a

matrix-valued sequence M_N, of *fixed* dimensions, such that $M_N/t_N \to 0$ when N tends to infinity.

Also, throughout this paper, we distinguish *Conditions* from *Assumptions*. Assumptions will refer to hypothesized properties of the system or its inputs; Assumptions may or may not hold. In contrast, Conditions can be enforced; enforcing them will be typically part of the process of designing the subspace algorithms.

A. From Hankel matrices to eigenstructure

Consider an observable pair (C, A) of matrices, where C is $q \times n$ and A is $n \times n$. Throughout this paper, p denotes an integer large enough such that

$$\text{rank}(\mathcal{O}_p) = n, \quad \text{where} \quad \mathcal{O}_p \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{p-1} \end{bmatrix} \tag{3}$$

For $i = 1, \ldots, p$ and $N > 0$, consider a family $R_i(N)$ of $q \times r$-matrices, satisfying the following assumption:

Condition 1: The matrices $R_i(N), N > 0$, decompose as $R_i(N) = CA^{i-1}G(N) + o(1)$. Furthermore, the sequence of $n \times r$-matrices $G(N), N > 0$, satisfies $\liminf_{N\to\infty} \sigma_n(G(N)) > 0$, where $\sigma_n(M)$ denotes the n-th largest singular value of matrix M.

Generic subspace algorithm: The following generic algorithm is considered throughout this paper. It assumes a finite family of matrices $R_i(N)$ as above and returns a pair $(C(N), A(N))$. The sentence

"$R_i(N)$ provides consistent estimators for (C, A)"

that we use throughout this paper means that, when provided with the sequence $R_i(N)$, this generic algorithm yields consistent estimators $(C(N), A(N))$ for the pair (C, A), see Theorem 1 below. Consider the matrix $\mathcal{H}_p(N)$ defined by

$$\mathcal{H}_p(N) \triangleq \begin{bmatrix} R_1(N) \\ R_2(N) \\ \vdots \\ R_p(N) \end{bmatrix} \tag{4}$$

and SVD-decompose it as:

$$\begin{aligned} \mathcal{H}_p(N) &= \textstyle\sum_{i=1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \\ &= \textstyle\sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^T + \sum_{i=n+1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \\ &= \mathbf{U}\, \text{diag}(\sigma_1, \ldots, \sigma_n)\, \mathbf{V}^T + \textstyle\sum_{i=n+1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \end{aligned}$$

Partition the above defined $pq \times n$ matrix $\mathbf{U}$ into its p successive q-block rows $\mathbf{U}_1, \ldots, \mathbf{U}_p$ and set

$$\mathbf{U}^{\uparrow} \triangleq \begin{bmatrix} \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_p \end{bmatrix} \quad \text{and} \quad \mathbf{U}^{\downarrow} \triangleq \begin{bmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_{p-1} \end{bmatrix}$$

Using these notations, set

$$\begin{aligned} C(N) &\triangleq \mathbf{U}_1 \\ A(N) &\triangleq \text{least-squares solution of } \mathbf{U}^{\uparrow} = \mathbf{U}^{\downarrow} A \end{aligned} \tag{5}$$

Theorem 1: [6] Under Condition 1, $(C(N), A(N))$ defined by (4–5) is a consistent estimator *of (C, A) in the following sense: there exists a sequence of matrices $T(N)$, with $T(N)$ and $T^{-1}(N)$ uniformly bounded w.r.t. N, such that $\lim_{N\to\infty} T^{-1}(N)A(N)T(N) \to A$, and $\lim_{N\to\infty} C(N)T(N) \to C$.*

Proof: It is found in [6], second part of Section III-C, dealing with the Balanced Realization algorithm. Besides the fact that reference [6] speaks (H, F, G) and not (A, B, C), the only slight change is that matrix $G(N)$ in Condition (1) replaces the controllability matrix $\mathbf{C}(F, G_S)$ of [6], where S is the sample length. ◇

B. Notations

For X and Y two matrices of compatible dimensions, define:

$$\begin{aligned} \langle X, Y \rangle &\triangleq XY^T \\ \|X\|^2 &\triangleq \text{Tr}\, \langle X, X \rangle \\ \text{E}(X \mid Y) &\triangleq \langle X, Y \rangle \langle Y, Y \rangle^{\dagger} Y \\ \text{E}(X \mid Y^{\perp}) &\triangleq X - \text{E}(X \mid Y)\ , \end{aligned} \tag{6}$$

where Tr denotes the trace and superscript † denotes the pseudo inverse. For $(y_k)_{k\in\mathbb{Z}}$ a $\mathbb{R}^q$-valued data sequence and $N > 0$ a window length, define

$$Y_i(N) \triangleq \begin{bmatrix} y_{i+N-1} & \cdots & y_{i+1} & y_i \end{bmatrix}$$

and write simply Y_i if N is understood. For $(x_k)_{k\in\mathbb{Z}}$ and $(z_k)_{k\in\mathbb{Z}}$ two data sequences of compatible dimensions, we write:

$$\langle X_i, Z_j \rangle_N \triangleq \langle X_i(N), Z_j(N) \rangle\ ,$$

and

$$\text{E}_N(X_i \mid Z_j) \triangleq \text{E}(X_i(N) \mid Z_j(N))\ .$$

Finally, we shall make use of the following data Hankel matrices:

$$\mathcal{Y}^+_{i,M}(N) \triangleq \begin{bmatrix} Y_{i+M} \\ \vdots \\ Y_{i+2} \\ Y_{i+1} \end{bmatrix},\ \mathcal{Y}^-_{i,M}(N) \triangleq \begin{bmatrix} Y_i \\ Y_{i-1} \\ \vdots \\ Y_{i-M} \end{bmatrix},$$

and

$$\mathcal{Y}_{i,M}(N) \triangleq \begin{bmatrix} \mathcal{Y}^+_{i,M} \\ \\ \mathcal{Y}^-_{i,M} \end{bmatrix}.$$

The above notations are introduced because, depending on the considered algorithms, the data set is indexed as $y_N, \ldots, y_1$ (when only "future" data are needed), or $y_N, \ldots, y_1, y_0, \ldots, y_{-N}$ (when data are split into future and past). Many authors use rather $y_1, \ldots, y_N, y_{N+1}, \ldots, y_{2N}$, or variants thereof. Clearly, the difference is only notational. Also, we have taken identical index M in $\mathcal{Y}^+_{i,M}$ and $\mathcal{Y}^-_{i,M}$ when building $\mathcal{Y}_{i,M}$. Of course, we could take different indices M_+ and M_- whithout impairing the validity of what follows.

Finally in order to refer to the different algorithms in a systematic way in the sequel, we shall superscript the referred $R_i(N)$ with the index of the corresponding equation. For example,

$$R_i^{(12)}(N) \text{ denotes } R_i(N) \text{ as specified by (12).} \tag{7}$$

C. Instruments

In this section, we revisit the old concept of "instrument" and use it in our context. Unlike in Section II where our problem was stated, we do not distinguish here between observed and unobserved inputs. In the following system, vector ξ collects all inputs of the system considered throughout this section:

$$\left\{ \begin{array}{rcl} x_k &=& Ax_{k-1} + B'\xi_k \\ y_k &=& Cx_{k-1} + D'\xi_k \end{array} \right. \tag{8}$$

In (8), $k \in \mathbb{Z}$, x is the $\mathbb{R}^n$-valued state, ξ is the $\mathbb{R}^m$-valued input, and y is the $\mathbb{R}^q$-valued observed output. Fix a window length N. With the notations of Section III-B, system (8) rewrites as follows, for $i = 1, \ldots, p$:

$$\left\{ \begin{array}{rcl} X_i &=& AX_{i-1} + B'\Xi_i \\ Y_i &=& CX_{i-1} + D'\Xi_i \end{array} \right. \tag{9}$$

In the following lemma we introduce *instruments* as the key tool in our analysis:

Lemma 1: (instruments) Let $(z_k)_{k\in\mathbb{Z}}$ be an $\mathbb{R}^M$-valued data sequence and $(s_N)_{N>0}$ an $\mathbb{R}_+$-valued sequence such that

$$\text{for } j \in \{1, \ldots, i\} \ : \ \langle \Xi_j, Z_0 \rangle_N = o(s_N) \tag{10}$$

$$\liminf_{N\to\infty} \sigma_n \left(\frac{1}{s_N} \langle X_0, Z_0 \rangle_N \right) > 0 \tag{11}$$

Then,

$$R_i(N) \triangleq \frac{1}{s_N} \langle Y_i, Z_0 \rangle_N \tag{12}$$

satisfies Condition 1. In the sequel, we call instrument *a signal (z_k) satisfying (10) and (11) for some sequence s_N.*

Proof: The following decompositions hold, for $i > 0$:

$$y_{k+i} = CA^{i-1}x_k + \sum_{j=1}^{i-1} CA^{i-1-j}B'\xi_{k+j} + D'\xi_{k+i},$$

with the convention that $\sum_1^0 = 0$, and

$$\begin{array}{l} \sum_{k=0}^{N-1} y_{k+i} z_k^T = \\ CA^{i-1} \sum_{k=0}^{N-1} x_k z_k^T + \\ \sum_{j=1}^{i-1} CA^{i-1-j} \sum_{k=0}^{N-1} B'\xi_{k+j} z_k^T + \sum_{k=0}^{N-1} D'\xi_{k+i} z_k^T \end{array} \tag{13}$$

Equation (13) rewrites as follows:

$$\begin{array}{l} \langle Y_i, Z_0 \rangle_N \\ = CA^{i-1} \langle X_0, Z_0 \rangle_N + \\ \sum_{j=1}^{i-1} CA^{i-1-j} B' \langle \Xi_j, Z_0 \rangle_N + D' \langle \Xi_i, Z_0 \rangle_N \ , \end{array} \tag{14}$$

which proves that $R^{(12)}(N)$ satisfies Condition 1, thanks to (10) and (11). ◇

Lemma 1 and Theorem 1 together ensure that $R^{(12)}(N)$ provides consistent estimators for the pair (C, A) (see (7) for the notational convention used here).

Applying Lemma 1 to system (1) with its combined observed and unobserved inputs can be (tentatively) performed via the following substitutions:

$$\left[\begin{array}{c} B' \\ D' \end{array} \right] \xi_k = \left[\begin{array}{c} B \\ D \end{array} \right] u_k + \left[\begin{array}{c} v_k \\ w_k \end{array} \right] \tag{15}$$

Of course, if input ξ_k is observed, *i.e.,* $v_k = w_k = 0$ in (15), then one can chose instrument z_k in such a way that $\langle \Xi_j, Z_0 \rangle_N = 0$ exactly. This is no longer feasible if unobserved inputs exist, since Ξ_j is no longer observed in this case. Therefore, additional work is needed for analysing system (1) with its combined observed/unobserved inputs. The next section addresses this missing point.

D. A martingale argument

Let us discuss the key conditions allowing us to apply Lemma 1 and Theorem 1 to system (1), taking the unobserved inputs v and w into account.

Suppose first that there is no unobserved input disturbance, *i.e.,* $v = w = 0$ in (1). Then, the observed values for the U_i's can be used while constructing instrument z_k in Lemma 1, since these are known—this is exactly what happens in the so-called "deterministic" subspace methods [19], where $\Xi_j = U_j$ holds and $\langle U_j, Z_0 \rangle_N = 0$ is enforced. Note that no assumption of stochastic nature is required for this reasoning.

Next, consider the opposite case in which there is no observed input, *i.e.,* $B = D = 0$ in (1). Since input disturbances are not observed, the actual values of Ξ_j are unknown when applying Lemma 1 and therefore cannot be used while constructing the instrument z_k.

This problem, however, can be solved by using *stochastic* knowledge about unobserved input disturbances. To this end, we now introduce the needed probabilistic setting.

Lemma 2: Let $(v_k)_{k\geq 0}$ and $(z_k)_{k\geq 0}$ be two sequences of square integrable vector valued random variables defined over some probability space $(\Omega, \mathcal{G}, \mathbb{P})$ and let $(\mathcal{G}_k)_{k\geq 0}$ be an increasing family of sub-σ-algebras of $\mathcal{G}$ such that:

$$\begin{array}{r} \sup_{k\geq 0} \mathbb{E}\|v_k\|^2 \leq K < \infty \ , \\ \lim_{N\to\infty} \sum_{k=0}^N \|z_k\|^2 = +\infty \text{ w.p.1} \ , \\ v_k \text{ and } z_k \text{ are } \mathcal{G}_k\text{-measurable, and} \\ \mathbb{E}(v_k \mid \mathcal{G}_{k-1}) = 0 \ . \end{array} \tag{16}$$

Then $\lim_{N\to\infty} \frac{M_N}{\sum_{k=0}^N \|z_k\|^2} = 0$ holds w.p. 1, where $M_N \triangleq \sum_{k=j}^N v_k z_{k-j}^T$ for $j > 0$.

Nota: In formula (16), the conditional expectation $\mathbb{E}(. \mid \mathcal{G}_{k-1})$ should not be confused with our matrix projection operator $\mathrm{E}(. \mid .)$ in (6).

Proof: It is a mild variation of the argument of [6], Section III-A. We repeat it here for the sake of completeness. Since we can reason on each entry of matrix M_N separately, we can, without loss of generality, assume that v_k and z_k are both scalar signals. By the second condition of (16), we know that $(M_k)_{k\geq 0}$ is a square integrable scalar martingale w.r.t. $(\mathcal{G}_k)_{k\geq 0}$. By (16), we have $\mathbb{E}((M_k - M_{k-1})^2 \mid \mathcal{G}_{k-1}) = \mathbb{E}(v_k^2 \mid \mathcal{G}_{k-1}) z_{k-j}^2 = \mathbb{E}(v_k^2) z_{k-j}^2 \leq K z_{k-j}^2$. The proof is

then completed by using Theorem 2 below, which can be found in [13], [17]. $\diamond$

Theorem 2: ([13], [17]) Let $(M_k)_{k\geq 0}$ be a locally square integrable martingale w.r.t. $(\mathcal{G}_k)_{k\geq 0}$, such that $M_0 = 0$. Set

$$[M,M]_k = \sum_{l=1}^{k} \mathbb{E}((M_l - M_{l-1})^2 \mid \mathcal{G}_{l-1}) \ .$$

*Then, w.p.*1*, we have $\frac{M_k}{[M,M]_k} \rightarrow 0$ on the set $\{\lim_{k\to\infty}[M,M]_k = +\infty\}$, and $\lim_{k\to\infty} M_k$ exists and is finite on the set $\{\lim_{k\to\infty}[M,M]_k < +\infty\}$.*

E. Analysing covariance based algorithms

In this section we combine the results from Sections III-C and III-D to handle system (1) with its combined observed/unobserved inputs. We repeat again system (1) for convenience:

$$\left\{ \begin{array}{rcccccc} x_k & = & Ax_{k-1} & + & Bu_k & + & v_k \\ y_k & = & Cx_{k-1} & + & Du_k & + & w_k \end{array} \right. \tag{17}$$

where $k \in \mathbb{Z}$, y is the $\mathbb{R}^q$-valued observed output, x is the $\mathbb{R}^n$-valued state, u is the $\mathbb{R}^m$-valued observed input, (v,w) is an unobserved input disturbance.

To be able to use stochastic information on the unobserved inputs v, w we assume that all variables arising in system (17) are defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Available information is captured by the following σ-algebras:

$$\begin{array}{rcl} \mathcal{F}_k & \stackrel{\Delta}{=} & \underbrace{\sigma\left(u_j : j \in \mathbb{Z}\right)}_{\mathcal{F}^u} \vee \underbrace{\sigma\left(y_l, v_l, w_l : l \leq k\right)}_{\mathcal{F}_k^{y,v,w}} \\ \mathcal{F}_k^o & \stackrel{\Delta}{=} & \underbrace{\sigma\left(u_j : j \in \mathbb{Z}\right)}_{\mathcal{F}^u} \vee \underbrace{\sigma\left(y_l : l \leq k\right)}_{\mathcal{F}_k^y} \end{array}$$

σ-algebra $\mathcal{F}^u$ is the information provided by the entire observed input sample; σ-algebra $\mathcal{F}_k^{y,v,w}$ is the information provided by the unobserved inputs v and w and the output y up to time k; finally, σ-algebra $\mathcal{F}_k^y$ is the information provided by the only output y up to time k. Regarding the unobserved inputs, we assume the following:

Assumption 1: [unobserved inputs] Stochastic inputs v and w satisfy the following conditions:

$$\begin{array}{l} \sup_{k\geq 0} \mathbb{E}\left(\|v_k\|^2 + \|w_k\|^2\right) < \infty, \text{ and for } j > 0 \ : \\ \forall k \geq 0, \mathbb{E}\left(v_{k+j} \mid \mathcal{F}_k\right) = 0 \text{ and } \mathbb{E}\left(w_{k+j} \mid \mathcal{F}_k\right) = 0 \ . \end{array}$$

Note that these conditions do not request any kind of stationarity. Assumption 1 involves the joint distribution of v_k, w_k, and u_k. It is in particular satisfied when observed and unobserved inputs are independent. Besides Assumption 1, no condition is required on the statistics of the observed input u_k. Consider the following conditions:

Condition 2: (instruments) Instrument (z_k) satisfies the following conditions:

$$z_k \text{ is } \mathcal{F}_k^o\text{-measurable} \tag{18}$$

$$\lim_{N\to\infty} s_N = \infty, \text{ where } s_N \stackrel{\Delta}{=} \sum_{k=-M}^{N-1} \|z_k\|^2 \tag{19}$$

$$\left\langle \left[\begin{array}{c} B \\ D \end{array} \right] U_j, Z_0 \right\rangle_N = o(s_N) \text{ for } j > 0 \tag{20}$$

$$\liminf_{N\to\infty} \sigma_n\left(\frac{1}{s_N}\langle X_0, Z_0\rangle_N\right) > 0 \tag{21}$$

Property (18) guarantees that instrument z_k depends only on observed quantities. Integer $M \geq 0$ in (19) is a constant selected according to each particular instance of the family $R_i(N)$. Property (19) expresses that instrument (z_k) possesses sustained energy. The following theorem is our first main result. It provides the analysis of algorithms of the form (12), *i.e.,* covariance based ones.

Theorem 3: (covariance based subspace) Assume that Assumption 1 regarding unobserved inputs, and Condition 2 regarding instruments, are in force. Then, $R_i^{(12)}(N)$ satisfies Condition 1.

Therefore, $R_i^{(12)}(N)$ yields a consistent subspace algorithm, by Theorem 1.

Proof: Using the notations of Section III-C, system (17) writes as follows, for $i = 1, \ldots, p$:

$$\left\{ \begin{array}{rcccccc} X_i & = & AX_{i-1} & + & BU_i & + & V_i \\ Y_i & = & CX_{i-1} & + & DU_i & + & W_i \end{array} \right. \tag{22}$$

On the other hand, system (17) yields the following decomposition for $y_{k+i}, i > 0$ (we use the convention that $\sum_1^0 = 0$):

$$y_{k+i} \ = \ CA^{i-1}x_k + \textstyle\sum_{j=1}^{i-1} CA^{i-1-j}\widehat{v}_{k+j} + \widehat{w}_{k+i}$$

where $\widehat{v}_k \stackrel{\Delta}{=} Bu_k + v_k$ and $\widehat{w}_k \stackrel{\Delta}{=} Du_k + w_k$. Using the notations of Section III-C, this decomposition rewrites as follows, for $i > 0$:

$$Y_i \ = \ CA^{i-1}X_0 + \textstyle\sum_{j=1}^{i-1} CA^{i-1-j}\widehat{V}_j + \widehat{W}_i \tag{23}$$

where $\widehat{V}_i \stackrel{\Delta}{=} BU_i + V_i$ and $\widehat{W}_i \stackrel{\Delta}{=} DU_i + W_i$. Note that

$$\langle V_j, Z_0\rangle_N \ = \ \textstyle\sum_{k=0}^{N-1} v_{k+j} z_k^T, \tag{24}$$

and a similar formula holds with W_j instead of V_j. By (18) and (19) of Condition 2, instrument (z_k) satisfies (16) in Lemma 2. By Assumption 1, noises v_k and w_k satisfy (16) in Lemma 2, with $\mathcal{F}_k$ substituted for $\mathcal{G}_k$. Therefore Lemma 2 can be applied with $\mathcal{F}_k$ substituted for $\mathcal{G}_k$, which yields, for any $j \in \{1, \ldots, p\}$:

$$\left\langle \left[\begin{array}{c} V_j \\ W_j \end{array} \right], Z_0 \right\rangle_N = o(s_N) \tag{25}$$

Set

$$\begin{array}{rcl} \xi_k & \triangleq & \left[\begin{array}{c} B \\ D \end{array} \right] u_k + \left[\begin{array}{c} v_k \\ w_k \end{array} \right] \\ B' & \triangleq & [\ I_n \quad 0_q\ ] \\ D' & \triangleq & [\ 0_n \quad I_q\ ] \end{array}$$

where the subscripts $_n$ and $_q$ indicate the dimensions of the corresponding matrices. Using this change of notation allows us to rewrite system (17) in the form (8) used in Lemma 1.

Consider now Condition 2. Combining (25) with (20) shows that system (8) satisfies (10) in Lemma 1. On the other hand (11) in Lemma 1 is ensured by Property (21) of Condition 2. Therefore, by Lemma 1 we conclude that Condition 1 is satisfied. ◇

Remarks: 1/ In fact our method could accomodate as well additional "small" perturbations in system (17), *i.e.,* additional inputs μ_k and ν_k in state and observation equations respectively, such that $\frac{1}{s_N}\sum_{k=-M}^{N-1} \|\mu_k\|^2 + \|\nu_k\|^2 = o(1)$. Transient terms or leakage effects such as considered in [7], [8] are covered by these additional terms, and therefore do not impair nonstationary consistency.

2/ Projection based subspace methods, *i.e.,* methods of the form

$$R_i(N) \triangleq \frac{1}{s_N} \mathrm{E}_N(Y_i \mid Z_0) \tag{26}$$

are in fact more popular than covariance based ones, see [20]. They are often referred to as "data based" subspace methods. Unfortunately, these methods cannot be handled directly by Theorem 3. In fact, Theorem 1 itself does not apply. The reason for this is simple: $R_i^{(26)}(N)$ has dimensions $q \times dim(Z_0(N))$. So its dimensions vary with N and therefore Theorem 1 cannot apply. In the full paper [1] a weighting technique is used to overcome this difficulty.

IV. Analysis of some subspace algorithms

In this section we apply our toolbox of theorems and lemmas to sample subspace methods. To avoid boring notational adjustments, we keep our notational conventions and will therefore sometimes deviate from the original presentations in this respect.

Key conditions ensuring nonstationary consistency are Assumption 1 and Condition 2. Assumption 1 involves the unobserved inputs, we assume it to hold throughout this section and will not discuss it any further. In contrast, Condition 2 is a *design constraint on the selection of the instruments*: this is the key condition to be verified or enforced when analysing specific algorithms.

Regarding the details of Condition 2, we shall pay great attention to verifying that (18) and (20) are satisfied, as these conditions drive the choice of the instruments. Condition (21) amounts to requiring that the instrument is well correlated to the state. In contrast, we shall not discuss the satisfaction of condition (19); this condition just translates, for each particular algorithm, into corresponding conditions for the original system (17).

Finally, checking for consistency requires that proper normalization is applied. This is the very role of the scaling factor s_N. In practice the algorithms are applied with given sample length N, and then, scaling is just an irrelevant issue. Therefore, we shall ignore scaling in this section.

A. Output-only (OO) subspace algorithms

By definition, these algorithms assume $B = D = 0$ in (17). Therefore (20) in Condition 2 is trivially satisfied, thus we essentially need to check the measurability property (18).

Basic OO subspace algorithm: This is the simplest algorithm to analyse. Introduce the instrument

$$z_k \triangleq \left[\begin{array}{c} y_k \\ \vdots \\ y_{k-M} \end{array} \right] \tag{27}$$

and take

$$R_i(N) = \langle Y_i, Z_0 \rangle_N. \tag{28}$$

Instrument (27) satisfies (18) in Condition 2. Hence Theorem 3 applies and proves consistency of $R_i^{(28)}(N)$. Note that

$$\langle X_0, Z_0 \rangle_N = [\ F(N) \quad AF(N) \quad \dots \quad A^M F(N)\ ],$$

where $F(N) = \langle X_0, Y_0 \rangle_N$. Hence, (21) can be interpreted as y_k being "uniformly of order n".

Covariance driven OO subspace algorithm [6], [18], [2], [12]: This algorithm is a variation of the previous algorithm, it was however proposed earlier. It consists in computing, for $i = 1, \dots, p$:

$$\begin{array}{l} R_i(N) = [\ \hat{r}_i(N) \quad \hat{r}_{i+1}(N) \quad \cdots \quad \hat{r}_{i+M}(N)\ ] \\ \text{where } \hat{r}_j(N) = \langle Y_j, Y_0 \rangle_N \end{array} \tag{29}$$

With instrument z_k as in (27), we have

$$R_i^{(29)}(N) - \langle Y_i, Z_0 \rangle_N = [\ \delta\langle Y_i, Y_0 \rangle \quad \dots \quad \delta\langle Y_i, Y_{-M} \rangle\ ]$$

where $\delta\langle Y_i, Y_{-k} \rangle \triangleq \langle Y_i, Y_{-k} \rangle_N - \langle Y_{i+k}, Y_0 \rangle_N$ is such that

$$\|\delta\langle Y_i, Y_{-k} \rangle\| \leq 2\, s^*_{M,N},$$

where

$$s^*_{M,N} \triangleq \sup_{-M \leq j \leq N-M} \sum_{l=j}^{j+M} \|y_l\|^2 \ .$$

This implies

$$\|R_i^{(29)}(N) - \langle Y_i, Z_0 \rangle_N\| = o(s_N)\ , \tag{30}$$

provided that the following assumption holds:

Assumption 2: For M fixed, $s^*_{M,N} = o(s_N)$.

With Assumption 2, (30) holds and therefore instrument z_k defined in (27) satisfies Condition 2. By Theorem 3, we derive that $R_i^{(29)}(N)$ yields a consistent subspace algorithm.

B. Input-output (IO) subspace algorithms

We review some representative variants.

Covariance driven IO subspace algorithms with projection on the orthogonal of the input [12]: This algorithm consists in computing (cf. notations (6)):

$$\begin{aligned} \hat{r}_i(N) &= \langle Z_i, Z_0 \rangle_N, \text{ where } Z_i \stackrel{\Delta}{=} \mathrm{E}_N\big(Y_i \,\big|\, \mathcal{U}_{0,M}^{\perp}\big) \\ R_i(N) &= \left[\begin{array}{cccc} \hat{r}_i(N) & \hat{r}_{i+1}(N) & \dots & \hat{r}_{i+M}(N) \end{array} \right] \end{aligned} \quad (31)$$

First, note that $\langle Z_i, Z_0 \rangle_N = \langle Y_i, Z_0 \rangle_N$. The associated instrument z_k is therefore the sequence of the successive columns of matrix Z_0. Note that z_k is $\mathcal{F}_k^o$-measurable. The rest of the analysis of this algorithm proceeds as for $R_i^{(29)}(N)$. Property (21) can be seen as that z_k itself being "uniformly of order n".

Covariance driven subspace algorithm using projected past inputs and outputs as instruments [25]: Those methods encompass the methods also known as IVM, CVA, PO-MOESP and N4SID in their covariance form [25]. In this paper, we will focus on the unweighted IV related to $\mathcal{H}_p$ defined as

$$\mathcal{H}_p = \langle \mathcal{Y}_{0,p}^{+}, \mathcal{L}_{0,M}^{-} \rangle_N \ , \quad (32)$$

where $\mathcal{L}_{0,M}^{-}$ is defined by stacking for $i = -M, \dots, 0$

$$L_i \stackrel{\Delta}{=} \mathrm{E}_N\left(\left[\begin{array}{c} U_i \\ Y_i \end{array} \right] \,\middle|\, \left(\mathcal{U}_{0,M}^{+}\right)^{\perp} \right) \ . \quad (33)$$

The rest of the analysis of this algorithm proceeds as for (31).

C. Time– vs. frequency–domain.

For $(y_k)_{k \in \mathbb{Z}}$ an $\mathbb{R}^q$-valued data sequence and $N > 0$ a window length, the DFT of $Y_i(N)$, denoted by $\widehat{Y}_i(N)$, is equal to $\widehat{Y}_i(N) = Y_i(N)\,\Delta_N^q$, where Δ_N^q is the N-DFT matrix. Since matrix Δ_N^q is orthogonal, $\langle \widehat{X}, \widehat{Y} \rangle_N = \langle X, Y \rangle_N$, and $\mathrm{E}_N(\widehat{X} \mid \widehat{Y}) = \mathrm{E}_N(X \mid Y)\,\Delta_N^q$. Hence, Condition 2 can be considered equivalently in the time domain or in the frequency domain. Therefore, frequency domain subspace algorithms corresponding to [15], [16] behave exactly the same way as their time domain counterparts regarding nonstationary consistency.

V. Conclusion

We have revisited eigenstructure identification via subspace methods. We have shown that consistency still holds in case of nonstationary inputs (in fact, for "nonstationary zero part").

Martingale techniques were used to deal with unobserved inputs—for these, "deterministic" projections based on observed data cannot be used; they can however be supplemented by "stochastic" projections using conditional expectations.

Not surprisingly, transient and leakage effects are not an issue for nonstationary consistency. And the results equivalently apply to both time- and frequency-domain methods.

The simple techniques developed in this paper are not able of addressing projection based algorithms. The interested reader is referred to the full version [1], where all classes of subspace methods are encompassed.

ACKNOWLEDGEMENT: *Michèle Basseville is gratefully acknowledged for corrections and useful suggestions on an earlier version of the manuscript.*

References

[1] Extended version of this paper. Submitted for publication.

[2] M. Basseville, A. Benveniste, M. Goursat, L. Hermans, L. Mevel, H. Van der Auweraer: "Output-only subspace-based structural identification: from theory to industrial testing practice". *ASME Journal of Dynamic Systems Measurement and Control*, Special Issue on Identification of Mechanical Systems, vol.123, no 4, Dec. 2001, 668-676.

[3] Bauer D., Deistler M. and Scherrer W. On the impact of weighting matrices in subspace algorithms. *in Proceedings of IFAC international symposium on identification*. Santa Barbara, CA. 2000.

[4] Bauer D., M Jansson M.. Analysis of the asymptotic properties of the MOESP type of subspace algorithms. *Automatica*, 36. 497-509. 2000.

[5] Bauer D., Deistler M. and Scherrer W. Consistency and asymptotic Normality of some subspace algorithms without observed inputs. *Automatica*, 35, 1243-1254, 2002.

[6] Benveniste, A. and Fuchs, J-J. Single sample modal identification of a nonstationary stochastic process. *IEEE Trans. on Automatic Control,* AC-30(1), 66-74, Jan. 1985.

[7] Cauberghe B. Applied frequency domain system identification in the field of experimental and operational modal analysis. PHD Thesis. Vrije Universitet Brussel, Belgium. 2004.

[8] Cauberghe B., Guillaume P. and Pintelon R. Combined Deterministic-Stochastic Frequency Domain Subspace. Submitted for publication.

[9] Chiuso A. and Picci G. Subspace identification by data orthogonalization and model decoupling. *Automatica*, 40, 1689-1704, 2004.

[10] Chiuso A. and Picci G. Asymptotic variance of Subspace methods by data orthogonalization and model decoupling: a comparative analysis *Automatica*, 40, 1705-1717, 2004.

[11] Deistler M, Peternell K. and Scherrer W. Consistency and relative efficiency of subspace methods. *Automatica*, 31, 1865-1875, 1995.

[12] Guyader A and Mevel L. Covariance driven subspace methods : input/output vs. output-only. *Proceedings of International Modal Analysis conference*, Orlando, February 2003.

[13] Hall P., and Heyde C.C. *Martingale limit theory and its applications.* Academic Press, Inc. (1980).

[14] Larimore W.E. System Identification, reduced order filters and modelling via canonical variate analysis. In *Proceedings of the American Control Conference*, 445-451, San Francisco, June 1983.

[15] Mc Kelvey T. Frequency domain system identification with instrumental based algorithm. In *Proc of the 16th biennal conference on mechanical vibration and noise*, Sacramento, Sept. 1997.

[16] Mc Kelvey T, Akcay H. and Ljung L. Subspace based multivariable system identification from frequency response data. *IEEE Transactions on Automatic Control*. 41-7, 960-978, 1996.

[17] Neveu J. *Discrete-parameter martingales.* Elsevier. 1975.

[18] Prevosto M., Olagnon M., Benveniste A., Basseville M., Le Vey G. State-space formulation, a solution to modal parameter estimation. *Journal of Sound and Vibration*. 148, no 2, pp.329-342, 1991.

[19] Van Overschee P., DeMoor B. N4Sid : Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30, 75-93, 1994.

[20] Van Overschee P., DeMoor B. *Subspace identification for linear systems.* Kluwer Academy Publishers. 1996.

[21] Verdult V., Verhaegen M. Subspace Identification of multivariable linear parameter-varying systems. *Automatica*, 38, 805-814, 2002.

[22] Verhaegen M. and Dewilde P. Subspace model identification part 3. Analysis of the ordinary output-error state space model identification algorithm. *International Journal of control*. 55, 555-586. 1993.

[23] Verhaegen M. Identification of the deterministic part of MIMO state space models given in innovation forms from input-output data. *Automatica*, 30, 61-74, 1994.

[24] Viberg M. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, 31(12), 1835-1851. 1995.

[25] Viberg M., Wahlberg, B. Ottersen B. Analysis of state space system identification methods based on instrumental variables and subspace fitting. *Automatica*, 33, 1603-1616. 1997.