Nonstationary Consistency of Subspace Methods

Albert Benveniste, *Fellow, IEEE*, 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 in the excitation and measurement noises. Note that such nonstationarities may result in having time-varying zeros for the underlying system, so the problem is nontrivial. In particular, likelihood- and prediction-error related methods do not ensure consistency under such situation, because estimation of poles and estimation of zeros are tightly coupled. We show in turn that subspace methods ensure such consistency. Our study carefully separates statistical from nonstatistical arguments, therefore, enlightening the role of statistical assumptions in this story.**

***Index Terms*—Consistency, nonstationary excitation, subspace methods.**

I. Introduction

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 in the excitation and measurement noises. Note that such nonstationarities may result in having time-varying zeros for the underlying system, so the problem is nontrivial. In particular, likelihood- and prediction-error-related methods do not ensure consistency under such situation, because estimation of poles and estimation of zeros are tightly coupled.

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 the class of nonstationary systems we discuss here. Since this paper, the family of subspace algorithms has been invented [16], [22], [25]–[27] and has expanded rapidly. Therefore, we felt it was timely revisiting the results of [6] and generalizing them to subspace methods. To this end, [6] had first to be restructured to show an important intermediate result, which had not been noticed explicitly in the original paper but was clearly there. Still, the generalization we present here is far less trivial than expected and required introducing new techniques for the proof.

There are a number of convergence studies on subspace methods in a stationary context in the literature; see [2]–[4], [10], [11], and [13], 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 the 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 estimation of the eigenstructure (the pole part) only and does not apply to the zero part, at least as far as the transfer from unobserved inputs to output measurements is concerned. It is definitely different from the problem considered in [24].

The paper is organized as follows. The problem of nonstationary consistency is stated in Section II, where a generic form of subspace algorithm is also stated. Section III collects the key steps of our analysis; Section III-A collects the nonprobabilistic arguments of the consistency proof; probabilistic arguments of the proof are collected in Section III-B; and our assumptions are discussed in Section III-C. Finally, in Section IV, by using the so developed toolbox of theorems and lemmas, we prove nonstationary consistency of some representative subspace algorithms.

II. Problem Setting—A Generic Subspace Algorithm

A. Problem Setting

Consider the following linear system:

$$\begin{cases} x_k = Ax_{k-1} + Bu_k + v_k \\ y_k = Cx_{k-1} + Du_k + w_k \end{cases} \tag{1}$$

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:

$$\begin{cases} x_k = Ax_{k-1} + Bu_k + K(k)\nu_k \\ y_k = Cx_{k-1} + Du_k + L(k)\nu_k \end{cases} \tag{2}$$

where

$$\begin{bmatrix} K(k) \\ L(k) \end{bmatrix} [\, K^T(k) \quad L^T(k) \,]$$

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 $\nu_k \mapsto y_k$ is time-varying in this case, so that consistency makes sense only with respect to (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

Manuscript received February 16, 2005; revised November 21, 2005, June 17, 2006, and October 22, 2006. Recommended by Associate Editor W. X. Zheng. This work was supported by the FliTE2 Eureka Project 2419.

The authors are with INRIA, Campus de Beaulieu, 35042 Rennes, France (e-mail: lmevel@inria.fr).

Digital Object Identifier 10.1109/TAC.2007.898970

φ_λ are a corresponding set of eigenvectors. Said in words, we consider the problem of *eigenstructure identification*.[1]

Our objective is to show that subspace methods provide consistent estimators of the eigenstructure, also for nonstationary cases as aforementioned. 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.

Now, we introduce the generic subspace algorithm we will analyze throughout this paper. This generic algorithm will be subsequently specialized to cover the various algorithms used in practice.

B. Generic Subspace Algorithm

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

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

Our generic algorithm assumes a finite family $R_i(N)$ of $q \times r$-matrices, where $r \geq n$, $i = 1, \ldots, p$ and $N > 0$. It returns a pair $(C(N), A(N))$. We describe it next. 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 decomposed by singular value decomposition (SVD) as

$$\begin{aligned} \mathcal{H}_p(N) &= \sum_{i=1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \\ &= \sum_{i=1}^{n} \sigma_i \mathbf{u}_i \mathbf{v}_i^T + \sum_{i=n+1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \\ &= \mathbf{U} \operatorname{diag}(\sigma_1, \ldots, \sigma_n) \mathbf{V}^T + \sum_{i=n+1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \\ &= \mathbf{U} S \mathbf{V}^T + \sum_{i=n+1}^{\min(pq,r)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \end{aligned} \tag{5}$$

Partition the $pq \times n$ matrix $\mathbf{U}$ defined in (5) 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

$$C(N) \triangleq \mathbf{U}_1 \tag{6}$$

$$A(N) \triangleq \text{least squares solutions of } \mathbf{U}^{\uparrow} = \mathbf{U}^{\downarrow} A. \tag{7}$$

Formulas (4)–(7) constitute our generic subspace algorithm. The remainder of the paper consists in analyzing this algorithm and specializations thereof. 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) in the sense made precise in Theorem 1.

III. Basic Theorems for Nonstationary Consistency

Throughout this paper, for t_N, a nondecreasing sequence of positive real numbers, $o(t_N)$ 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 satisfied by proper design of the algorithm; enforcing these Conditions will be typically part of the process of designing the subspace algorithms.

Our analysis proceeds in two steps. The first step collects the arguments that do not involve probability, whereas only the second step makes use of statistical arguments.

A. Nonprobabilistic Analysis

In this section, we collect all arguments of the analysis that make no use of probability at all. Therefore, "convergence" is meant here in the usual, nonprobabilistic, sense.

1) From Hankel Matrices to Eigenstructure: For $i = 1, \ldots, p$ and $N > 0$, consider a family $R_i(N)$ of $q \times r$-matrices, satisfying Condition 1.

Condition 1: The matrices $R_i(N), N > 0$, decompose as

$$R_i(N) = CA^{i-1}G(N) + o(1). \tag{8}$$

Furthermore, the sequence of $n \times r$-matrices $G(N), N > 0$, satisfies the following condition:

$$\liminf_{N \to \infty} \sigma_n(G(N)) > 0 \tag{9}$$

where $\sigma_n(M)$ denotes the nth largest singular value of matrix M.

Theorem 1 (Consistent Estimator [6]): Under Condition 1, $(C(N), A(N))$ defined by (4)–(7) 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, Sec. III-C], dealing with the balanced realization algorithm. Besides the fact that [6] speaks

[1] This problem and the situation described in (2) naturally occur, for example, in the modal analysis of mechanical structures subject to vibration under both controlled and/or natural and turbulent excitation [1].

(H, F, G) and not (A, B, C), the only slight change is that matrix $G(N)$ in (8) replaces the controllability matrix $\mathbf{C}(F, G_S)$ of [6], where S is the sample length.

In the following, we will relate our matrices $R_i(N)$ to empirical covariances of data. For this, we need some more notations.

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

$$\begin{aligned} \langle X, Y\rangle &\triangleq XY^T \\ \|X\|^2 &\triangleq \mathrm{Tr}\,\langle X, X\rangle \\ \mathrm{E}(X \,|\, Y) &\triangleq \langle X, Y\rangle\langle Y, Y\rangle^\dagger Y \\ \mathrm{E}(X \,|\, Y^\perp) &\triangleq X - \mathrm{E}(X \,|\, Y) \end{aligned} \tag{10}$$

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

$$Y_i(N) \triangleq [\, y_{i+N-1} \quad \cdots \quad y_{i+1} \quad y_i \,]$$

and write simply Y_i if no confusion can result. 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

$$\mathrm{E}_N(X_i \,|\, Z_j) \triangleq \mathrm{E}(X_i(N) \,|\, Z_j(N)).$$

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

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

and

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

The previous 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_- without impairing the validity of what follows.

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

$$R_i^{(16)}(N) \text{ denotes } R_i(N) \text{ as specified by (16).} \tag{11}$$

The same convention will be used when we wish to refer to algorithms in terms of their $\mathcal{H}$ matrix.

3) 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

$$\begin{cases} x_k = Ax_{k-1} + B'\xi_k \\ y_k = Cx_{k-1} + D'\xi_k \end{cases}. \tag{12}$$

In (12), $k \in \mathbb{Z}$, 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-A2, system (12) rewrites as follows, for $i = 1, \ldots, p$:

$$\begin{cases} X_i = AX_{i-1} + B'\Xi_i \\ Y_i = CX_{i-1} + D'\Xi_i \end{cases}. \tag{13}$$

In Lemma 1, 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

$$\text{for } j \in \{1, \ldots, i\} : \langle \Xi_j, Z_0\rangle_N = o(s_N) \tag{14}$$

$$\liminf_{N\to\infty} \sigma_n\left(\frac{1}{s_N}\langle X_0, Z_0\rangle_N\right) > 0. \tag{15}$$

Then

$$R_i(N) \triangleq \frac{1}{s_N}\langle Y_i, Z_0\rangle_N \tag{16}$$

satisfies Condition 1. In the sequel, we call instrument a signal (z_k) satisfying (14) and (15) 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

$$\begin{aligned} \sum_{k=0}^{N-1} y_{k+i}z_k^T = {} & CA^{i-1}\sum_{k=0}^{N-1} x_k z_k^T \\ & + \sum_{j=1}^{i-1} CA^{i-1-j}\sum_{k=0}^{N-1} B'\xi_{k+j}z_k^T \\ & + \sum_{k=0}^{N-1} D'\xi_{k+i}z_k^T. \end{aligned} \tag{17}$$

Equation (17) rewrites as follows:

$$\begin{aligned} \langle Y_i, Z_0\rangle_N = {} & CA^{i-1}\langle X_0, Z_0\rangle_N \\ & + \sum_{j=1}^{i-1} CA^{i-1-j}B'\langle \Xi_j, Z_0\rangle_N \\ & + D'\langle \Xi_i, Z_0\rangle_N \end{aligned} \tag{18}$$

which proves that $R^{(16)}(N)$ satisfies Condition 1, thanks to (14) and (15).

Lemma 1 and Theorem 1 together ensure that $R^{(16)}(N)$ provides consistent estimators for the pair (C, A)—see (11) 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:

$$\begin{bmatrix} B' \\ D' \end{bmatrix} \xi_k = \begin{bmatrix} B \\ D \end{bmatrix} u_k + \begin{bmatrix} v_k \\ w_k \end{bmatrix}. \tag{19}$$

Of course, if input ξ_k is observed, i.e., $v_k = w_k = 0$ in (19), 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 analyzing system (1) with its combined observed/unobserved inputs. Section III-B on probabilistic analysis will address this missing point.

4) Weighting and Squaring: (This section may be ignored for a first reading.)

As perfectly analyzed in [23], there are many different subspace algorithms, and, in addition, each of these possesses a number of variants. Such variants depend on whether the algorithm uses raw data or frequency domain spectra, or time-domain covariance matrices as inputs; they also depend on which type of "weighting" is being used. In this section, we will develop a toolbox of lemmas to show that, once one of these variants is shown to be consistent, then so are all related variants. Our toolbox involves the following two tools: weighting and squaring.

a) Weighting: Weighting is generally used as part of subspace algorithms and plays an important role in algorithm conditioning and convergence rates. In our case, weighting will be in addition a key tool for the analysis of algorithms, should they be weighted or not. In particular, some subspace algorithms take as input matrices $R_i(N)$ whose dimensions are not fixed but vary with the length N of the data set. Consequently, the apparatus of Section II does not apply directly to such matrices $R_i(N)$. Weighting will be used as a preliminary step in analyzing such algorithms.

We distinguish preweighting, indicated by the symbol λ in sub- or superscript, and postweighting, indicated by the symbol ρ in sub- or superscript. Symbols λ and ρ are reminiscent of "left" and "right," respectively. Preweighting consists in premultiplying the matrix $\mathcal{H}_p$ defined in (4) by a square and invertible matrix W_λ. Postweighting consists in postmultiplying $\mathcal{H}_p$ by a rectangular matrix W_ρ^T, of dimensions possibly varying with the length N of the record. In this discussion, we omit index N when no ambiguity can result.

In what follows, superscript ^{w} attached to R_i or $\mathcal{H}_p$ announces that corresponding matrices cannot be handled directly by the apparatus of Section II, and thus, weighting will be used in analyzing the corresponding algorithm.

Let r_N be a sequence of positive integers (indexed by the length N of the data set). We are given the following:

- — a family $R_i^w(N)$ of $q \times r_N$-matrices, where $i = 1, \ldots, p$;
- — a sequence $W_\lambda(N)$ of preweighting matrices of dimensions $pq \times pq$;
- — a sequence $W_\rho^T(N)$ of postweighting matrices of dimensions $r_N \times r$.

Let $\mathcal{H}_p^w(N)$ be the matrix obtained by stacking the matrices $R_i^w(N)$ as in (4). Then, set $\mathcal{H}_p(N) = W_\lambda(N)\,\mathcal{H}_p^w(N)\,W_\rho^T(N)$. Partitioning $\mathcal{H}_p(N)$ as in (4) defines a family $R_i(N)$ of matrices. Now, SVD-decomposing $\mathcal{H}_p(N)$ yields

$$\mathcal{H}_p(N) = \mathbf{U}\mathrm{diag}(\sigma_1, \ldots, \sigma_n)\mathbf{V}^T + \sum_{i=n+1}^{\min(pq,r)} \sigma_i \mathbf{u}_i {\mathbf{v}_i}^T. \tag{20}$$

For given N, let $(C(N), A(N))$ be the pair obtained by applying formulas (6) and (7) to the matrix $\mathbf{U}$. On the other hand, SVD-decompose $\mathcal{H}_p^w(N)$ as

$$\mathcal{H}_p^w(N) = \mathbf{U}_w \mathrm{diag}\,(\sigma_1^w, \ldots, \sigma_n^w)\,\mathbf{V}_w^T + \sum_{i=n+1}^{\min(pq,r)} \sigma_i^w \mathbf{u}_i^w \mathbf{v}_i^{wT} \tag{21}$$

and set $\mathbf{U}' = W_\lambda \mathbf{U}_w$. Then, let $(C_w(N), A_w(N))$ be the pair obtained by applying formulas (6) and (7) to the matrix $\mathbf{U}'$.

Note that the family $R_i(N)$ possesses constant dimensions and is, therefore, amenable to a direct application of Theorem 1. In contrast, the family $R_i^w(N)$ cannot satisfy Condition 1 since its dimensions are $q \times r_N$ and thus may vary with N. Therefore, the consistency of $(C_w(N), A_w(N))$ cannot follow from a direct application of Theorem 1.

Lemma 2 overcomes this difficulty by making it possible to transfer consistency from $(C(N), A(N))$ to $(C_w(N), A_w(N))$.

To this end, note that pre- and postmultiplying (21) by $W_\lambda(N)$ and $W_\rho^T(N)$ yields

$$\mathcal{H}_p(N) = W_\lambda(N)\mathbf{U}_w \mathrm{diag}\,(\sigma_1^w, \ldots, \sigma_n^w)\,\mathbf{V}_w^T W_\rho^T(N) + W_\lambda(N) \left(\sum_{i=n+1}^{\min(pq,r)} \sigma_i^w \mathbf{u}_i^w \mathbf{v}_i^{wT} \right) W_\rho^T(N). \tag{22}$$

Lemma 2 (Weighting): Assume that the sequence $\mathcal{H}_p(N)$ is bounded w.r.t. N and that the following condition holds:

$$\limsup_{N \to \infty} W_\lambda(N) \left(\sum_{i=n+1}^{\min(pq,r)} \sigma_i^w \mathbf{u}_i^w \mathbf{v}_i^{wT} \right) W_\rho^T(N) = 0. \tag{23}$$

Then, the pair $(C_w(N), A_w(N))$ is consistent iff the pair $(C(N), A(N))$ is consistent.

Proof: See the Appendix.

b) Squaring: Squaring is a particular case of postweighting, where the weighting matrix is just the transpose of the original one. Squaring is an instrumental tool in analyzing projection-based algorithms.

Corollary 1 (Squaring): With the same notations as before, assume that $\mathcal{H}_p(N)$ and $\mathcal{H}_p^w(N)$ are related by $\mathcal{H}_p(N) = \mathcal{H}_p^w(N)\mathcal{H}_p^w(N)^T$.

1) If $\mathcal{H}_p(N)$ satisfies Condition 1, then the pair $(C_w(N), A_w(N))$ is consistent.
2) Vice-versa, if $\mathcal{H}_p^w(N)$ satisfies Condition 1, then the pair $(C(N), A(N))$ is consistent.

Proof: See the Appendix.

B. Probabilistic Analysis

So far, probabilities were never invoked. In this section, we collect the arguments involving probability and assumptions of probabilistic nature.

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 ξ_ks introduced in (12) are observed and thus can be explicitly used to satisfy a stronger condition than (14) in Lemma 1, namely $\langle \Xi_j, Z_0 \rangle_N = 0$. 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, and, prior to this, the martingale argument we will use.

1) Martingale Argument:

Lemma 3: 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

$$\sup_{k\geq 0} \mathbb{E}\|v_k\|^2 \leq \mathbf{C} < \infty$$

and

$$\lim_{N\to\infty} \sum_{k=0}^{N} \|z_k\|^2 = +\infty \text{ w.p. } 1 \tag{24}$$

v_k and z_k are $\mathcal{G}_k$-measurable and $\mathbb{E}(v_k \,|\, \mathcal{G}_{k-1}) = 0$. Then, for any $j > 0$, the following holds:

$$\lim_{N\to\infty} \frac{M_N(j)}{\sum_{k=0}^{N} \|z_k\|^2} = 0 \text{ w.p. } 1,$$

$$\text{where} \quad M_N(j) \triangleq \sum_{k=j}^{N} v_k z_{k-j}^T. \tag{25}$$

Notice: In (25), the conditional expectation $\mathbb{E}(. \,|\, \mathcal{G}_{k-1})$ should not be confused with our matrix projection operator $\mathrm{E}(.|.)$ in (10).

Proof: It is a mild variation of the argument of [6, Sec. 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 (25), we know that $(M_k)_{k\geq 0}$ is a square integrable scalar martingale w.r.t. $(\mathcal{G}_k)_{k\geq 0}$. By (25), we have $\mathbb{E}((M_k - M_{k-1})^2 \,|\, \mathcal{G}_{k-1}) = \mathbb{E}(v_k^2 \,|\, \mathcal{G}_{k-1}) z_{k-j}^2 = \mathbb{E}(v_k^2) z_{k-j}^2 \leq \mathbf{C} z_{k-j}^2$. The proof is then completed by using Theorem 2, which can be found in [15] and [20].

The real-valued stochastic process $(M_k)_{k\geq 0}$ is called a locally square integrable martingale w.r.t. $(\mathcal{G}_k)_{k\geq 0}$ if, for every $k, \mathbb{E}(M_k \,|\, \mathcal{G}_{k-1}) = 0$, and, for every $K < \infty, \sup_{k\leq K} \mathbb{E} M_k^2 < \infty$.

Theorem 2 [15], [20]: 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

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

Then

$$\frac{M_k}{[M, M]_k} \to 0 \text{ w.p. } 1$$

on the set where $\lim_{k\to\infty} [M, M]_k = +\infty$ holds. On the other hand

$$\lim_{k\to\infty} M_k \text{ exists and is finite w.p. } 1$$

on the set where $\lim_{k\to\infty} [M, M]_k < +\infty$ holds.

2) Analyzing the Generic Subspace Algorithm: In this section, we combine the results from Sections III-A3 and III-B1 to handle system (1) with its combined observed/unobserved inputs. We repeat again system (1) for convenience

$$\begin{cases} x_k = Ax_{k-1} + Bu_k + v_k \\ y_k = Cx_{k-1} + Du_k + w_k \end{cases} \tag{26}$$

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, and (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 (26) are defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Available information is captured by the following σ-algebras:

$$\mathcal{F}_k \triangleq \underbrace{\sigma(u_j : j \in \mathbb{Z})}_{\mathcal{F}^u} \vee \underbrace{\sigma(y_l, v_l, w_l : l \leq k)}_{\mathcal{F}_k^{y,v,w}}$$

$$\mathcal{F}_k^o \triangleq \underbrace{\sigma(u_j : j \in \mathbb{Z})}_{\mathcal{F}^u} \vee \underbrace{\sigma(y_l : l \leq k)}_{\mathcal{F}_k^y}$$

where $\mathcal{F} \vee \mathcal{F}'$ denotes the smallest σ-algebra containing $\mathcal{F}$ and $\mathcal{F}'$. In these formulas, 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 inputs, we assume the following.

Assumption 1 (Regarding Inputs): Stochastic inputs v and w satisfy the following conditions:

$$\sup_{k\geq 0} \mathbb{E}(\|v_k\|^2 + \|w_k\|^2) < \infty$$

$$\forall j > 0 \qquad \forall k \geq 0 : \mathbb{E}(v_{k+j} \,|\, \mathcal{F}_k) = 0 \text{ and } \mathbb{E}(w_{k+j} \,|\, \mathcal{F}_k) = 0.$$

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 regarding instruments.

Condition 2 (Regarding Instruments): Instrument (z_k) satisfies the following conditions:

$$z_k \text{ is } \mathcal{F}_k^o\text{−measurable} \tag{27}$$

$$\lim_{N\to\infty} s_N = \infty, \qquad \text{where} \quad s_N \triangleq \sum_{k=-M}^{N-1} \|z_k\|^2 \tag{28}$$

$$\left\langle \begin{bmatrix} B \\ D \end{bmatrix} U_j, Z_0 \right\rangle_N = o(s_N) \text{ for } j > 0 \tag{29}$$

$$\liminf_{N\to\infty} \sigma_n \left(\frac{1}{s_N} \langle X_0, Z_0 \rangle_N \right) > 0. \tag{30}$$

Property (27) guarantees that instrument z_k depends only on observed quantities. Integer $M \geq 0$ in (28) is a constant selected according to each particular instance of the family $R_i(N)$. Property (28) expresses that instrument (z_k) possesses sustained energy.

a) Covariance-based subspace: Theorem 3 is our first main result. It provides the analysis of algorithms of the form (16), 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, $R_i^{(16)}(N)$ satisfies Condition 1, with Probability 1.

In other words, the set of trajectories of the system for which Condition 1 is satisfied has Probability 1. Pick any trajectory in this set; we can apply Theorem 1, which shows that, for this trajectory, our generic algorithm provides consistent estimators in the sense of Theorem 1. This shows that our generic algorithm provides consistent estimators in the statistical sense (convergence w.p. 1 to the true value for the parameters to be estimated).

Proof: Using the notations of Section III-A2, system (26) writes as follows, for $i = 1, \ldots, p$:

$$\begin{cases} X_i = AX_{i-1} + BU_i + V_i \\ Y_i = CX_{i-1} + DU_i + W_i \end{cases}. \tag{31}$$

On the other hand, system (26) yields the following decomposition for $y_{k+i}, i > 0$ (we use the convention that $\sum_1^0 = 0$):

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

where $\hat{v}_k \triangleq Bu_k + v_k$ and $\hat{w}_k \triangleq Du_k + w_k$. Using the notations of Section III-A3, this decomposition rewrites as follows, for $i > 0$:

$$Y_i = CA^{i-1}X_0 + \sum_{j=1}^{i-1} CA^{i-1-j}\hat{V}_j + \hat{W}_i \tag{32}$$

where $\hat{V}_i \triangleq BU_i + V_i$ and $\hat{W}_i \triangleq DU_i + W_i$. Note that

$$\langle V_j, Z_0 \rangle_N = \sum_{k=0}^{N-1} v_{k+j} z_k^T \tag{33}$$

and a similar formula holds with W_j instead of V_j. By (27) and (28) of Condition 2, instrument (z_k) satisfies (25) in Lemma 3. By Assumption 1, noises v_k and w_k satisfy (25) in Lemma 3, with $\mathcal{F}_k$ substituted for $\mathcal{G}_k$. Therefore, Lemma 3 can be applied with $\mathcal{F}_k$ substituted for $\mathcal{G}_k$, which yields, with Probability 1

$$\forall j \in \{1, \ldots, p\} : \left\langle \begin{bmatrix} V_j \\ W_j \end{bmatrix}, Z_0 \right\rangle_N = o(s_N). \tag{34}$$

Set

$$\xi_k \triangleq \begin{bmatrix} B \\ D \end{bmatrix} u_k + \begin{bmatrix} v_k \\ w_k \end{bmatrix}$$
$$B' \triangleq [\, I_n \quad 0_q \,]$$
$$D' \triangleq [\, 0_n \quad I_q \,]$$

where the subscripts $_n$ and $_q$ indicate the dimensions of the corresponding matrices. Using this change of notation allows us to rewrite system (26) in the form (12) used in Lemma 1.

Consider now Condition 2. Combining (34) with (29) shows that system (12) satisfies (14) in Lemma 1. On the other hand, (15) in Lemma 1 is ensured by (30) of Condition 2. Therefore, by Lemma 1, we conclude that Condition 1 is satisfied, with Probability 1.

Remark: In fact, our method could accommodate as well additional "small" perturbations in system (26), 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 [8] and [9] are covered by these additional terms, and therefore, do not impair nonstationary consistency.

b) Projection-based subspace: Projection-based subspace methods, i.e., methods of the form

$$R_i(N) \triangleq \frac{1}{s_N} \mathrm{E}_N(Y_i \,|\, Z_0) \tag{35}$$

are in fact more popular than covariance-based ones; see [23]. 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: $R_i^{(35)}(N)$ has dimensions $q \times \dim(Z_0(N))$; so its dimensions vary with N, and therefore, Theorem 1 cannot apply. Fortunately, the weighting technique of Section III-A4 can be used to overcome this difficulty as we will see now.

Corollary 1 of Section III-A4 can be used to relate covariance-based methods, i.e., of the form

$$R_i'(N) = \frac{1}{s_N} \langle Y_i, Z_0 \rangle_N$$

to projections-based ones

$$R_i(N) = \frac{1}{s_N} \mathrm{E}_N(Y_i \,|\, Z_0). \tag{36}$$

The former are handled by Theorem 3 but the latter are not. To establish this relation, stack the matrices $R_i(N)$ as usual and consider

$$\mathcal{H}_p(N) = \mathrm{E}_N\left(\mathcal{Y}^+_{0,n} \,|\, Z_0\right). \tag{37}$$

Note that this Hankel matrix has its second dimension that varies with the length N of the data sample and thus cannot be handled by Theorem 3. To study this algorithm, we will, therefore, use a "squaring" method based on Corollary 1. Consider

$$\begin{aligned}\mathcal{K}_p(N) &= \mathcal{H}_p(N)\mathcal{H}_p(N)^T \\ &= \langle \mathcal{Y}^+_{0,p}, Z_0\rangle_N \langle Z_0, Z_0\rangle_N^{\dagger} \langle Z_0, \mathcal{Y}^+_{0,p}\rangle_N.\end{aligned}$$

By point 1 of Corollary 1, it is enough to guarantee that $\mathcal{K}_p(N)$ satisfies Condition 1. To this end, renormalize instrument Z_0

$$\hat{Z}_0 = \langle Z_0, Z_0\rangle_N^{-\frac{1}{2}} Z_0(N) \tag{38}$$

where superscript $^{-1/2}$ denotes the square root of the pseudoinverse. Note that (38) amounts to whitening instrument z_k. Then, $\mathcal{K}_p(N)$ rewrites

$$\mathcal{K}_p(N) = \left\langle \mathcal{Y}^+_{0,p}, \hat{Z}_0 \right\rangle_N \left\langle \hat{Z}_0, \mathcal{Y}^+_{0,p} \right\rangle_N. \tag{39}$$

By point 2 of Corollary 1, it is enough to guarantee that the square root matrix $\langle \mathcal{Y}^+_{0,p}, \hat{Z}_0\rangle_N$ satisfies Condition 1. Note that instrument $\hat{Z}_0$ satisfies the measurability property (27) of Condition 2. Our second main result, therefore, states as follows.

Theorem 4 (Projection-Based Subspace): Assume that Assumption 1 is in force, as well as properties (27)–(30) of Condition 2, for instrument (38). Then, $R_i^{(36)}(N)$ satisfies Condition 1, with Probability 1.

Consequently, $R_i^{(36)}(N)$ yields a consistent subspace algorithm, in the statistical sense.

C. Discussing Assumptions and Conditions

Here, we collect remarks concerning our assumptions and conditions.

What if matrix A is unstable?: Strictly speaking, it is nowhere required that matrix A should be stable. However, if A has some unstable eigenvalues and some stable ones, then property (30) of Condition 2 can hardly be satisfied.

What Can the Observed Inputs u Be?: Property (29) of Condition 2 relates instrument z_k to input u_k; however, the latter condition should rather be seen as a condition on the instrument, not as a condition on the input. The only important requirement on u is Assumption 1. This assumption is, in particular, satisfied if future unobserved inputs $v_{k+j}, w_{k+j}, j > 0$ are independent from inputs $u_l, l \in \mathbb{Z}$. For example, if (v, w) is white noise, Assumption 1 is satisfied if u_k does not depend on future outputs.[2] On the other hand, there is no requirement per se that u should be stationary. In some sense, the probability distribution of u does not matter and we regard u as stochastic in Section III-B2 only for mathematical convenience.

[2]Note that the latter condition is not compatible with closed-loop identification.

What Is Really Allowed Regarding Unobserved Input Noise (v, w)?: First, the time-varying matrices $K(k)$ and $L(k)$ in (2) may be random. This must, however, occur in a way that Assumption 1 is not be invalidated. For example, referring to (2), it is possible that $K(k)$ and $L(k)$ are stochastic processes that are independent from both underlying white noise ν_k and observed input u_k.

Can (v, w) be colored in (1)?: Yes, in part. In fact, moving average measurement noise is allowed

$$w_k = \sum_{j=0}^{J} L_j \nu_{k-j}$$

where ν is a possibly nonstationary white noise. To see this, rewrite (1) as follows, with $\xi_k^T \triangleq [\nu_k^T, \ldots, \nu_{k-J+1}^T]$:

$$\left\{\begin{array}{rcccccc} x_k &=& Ax_{k-1} &+& Bu_k &+& v_k \\ \xi_k &=& S\xi_{k-1} &+& &+& T\nu_k \\ y_k &=& CL\begin{bmatrix} x_{k-1} \\ \xi_{k-1}\end{bmatrix} &+& Du_k &+& L_0\nu_k \end{array}\right. \tag{40}$$

where $CL = [C \quad \overline{L}]$, S is the nilpotent matrix having 1s on the lower diagonal entries and 0s elsewhere, $T = [I0\ldots0]^T$, and $\overline{L} = [L_1 \ldots L_J]$. Applying the generic algorithm with (16) to system (40) yields the eigenstructure of the pair

$$\left([C \quad \overline{L}], \begin{bmatrix} A & 0 \\ 0 & S\end{bmatrix}\right)$$

where we recall that S has 0 as unique (multiple) eigenvalue. When A does not have 0 as an eigenvalue, this immediately yields the desired eigenstructure of the pair (C, A).

IV. Analysis of Some Subspace Algorithms

In this section, we apply our toolbox of theorems and lemmas to sample subspace methods. To avoid annoying 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 analyzing specific algorithms.

Regarding the details of Condition 2, we will pay great attention to verifying that (27) and (29) are satisfied, as these conditions drive the choice of the instruments. Condition (30) amounts to requiring that the instrument is well correlated to the state. In contrast, we will not discuss the satisfaction of condition (28); this condition just translates, for each particular algorithm, into corresponding conditions for the original system (26).

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 will ignore scaling in this section.

A. Output-Only (OO) Subspace Algorithms

By definition, these algorithms assume $B = D = 0$ in (26). Therefore, (29) in Condition 2 is trivially satisfied for such algorithms.

a) Basic OO subspace algorithm: This is the simplest algorithm to analyze. Introduce the instrument

$$z_k \triangleq \begin{bmatrix} y_k \\ \vdots \\ y_{k-M} \end{bmatrix}, \qquad \text{or equivalently} \quad Z_0 = \begin{bmatrix} Y_0 \\ \vdots \\ Y_{-M} \end{bmatrix} \tag{41}$$

and take

$$R_i(N) = \langle Y_i, Z_0 \rangle_N. \tag{42}$$

Instrument (41) satisfies (27) in Condition 2. Points (27) and (29) are the key points of Condition 2. Regarding the other points, note that

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

where $F(N) = \langle X_0, Y_0 \rangle_N$; hence, (30) in Condition 2 can be interpreted as y_k being "uniformly of order n." Finally, (28) in Condition 2 is a mild condition related to excitation persistency. These two remaining points will not be discussed any further in the sequel and we will, therefore, focus on (27) and (29).

In conclusion, Theorem 3 applies and proves consistency of $R_i^{(42)}(N)$.

b) Covariance-driven OO subspace algorithm [1], [6], [14], [21]: This algorithm is a variation of the previous algorithm; it was, however, proposed earlier. It consists in computing, for $i = 1, \ldots, p$

$$R_i(N) = [\, \hat{r}_i(N) \quad \hat{r}_{i+1}(N) \quad \cdots \quad \hat{r}_{i+M}(N) \,], \qquad \text{where} \quad \hat{r}_j(N) = \langle Y_j, Y_0 \rangle_N. \tag{43}$$

With instrument z_k as in (41), we have

$$R_i^{(43)}(N) - \langle Y_i, Z_0 \rangle_N = [\, \delta\langle Y_i, Y_0 \rangle \quad \cdots \quad \delta\langle Y_i, Y_{-M} \rangle \,]$$

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

$$\left\| R_i^{(43)}(N) - \langle Y_i, Z_0 \rangle_N \right\| = o(s_N) \tag{44}$$

provided that the Assumption 2 holds.

Assumption 2: For M fixed, $s^*_{M,N} = o(s_N)$: Under the previous additional assumption, (44) holds, and therefore, instrument z_k defined in (41) satisfies Condition 2. Therefore, by Theorem 3, we derive that $R_i^{(43)}(N)$ yields a consistent subspace algorithm.

c) Data-driven OO subspace algorithms [23]: This algorithm is found in [23, Ch. 3, Th. 8]. It consists in computing

$$\mathcal{H}_p(N) = \mathrm{E}_N \left(\mathcal{Y}^+_{0,p} \,|\, \mathcal{Y}^-_{0,M} \right). \tag{45}$$

To study this algorithm, we will use Theorem 4 about projection-based methods. To this end, set

$$\hat{Z}_0 = \left(\left\langle \mathcal{Y}^-_{0,M}, \mathcal{Y}^-_{0,M} \right\rangle_N \right)^{-\frac{1}{2}} \mathcal{Y}^-_{0,M}(N) \tag{46}$$

which amounts to whitening the instrument (41). Instrument (46) satisfies the measurability condition (27) of Condition 2. Assuming that (28) and (30) are satisfied, this yields consistency, by Theorem 4.

B. Input–Output (IO) Subspace Algorithms

Many variants have been considered. We review some representative ones.

a) Covariance-driven subspace algorithm using projected past inputs and outputs as instruments [28]: Those methods encompass the methods also known as instrumental variable method (IVM), canonical variate analysis (CVA), past output multivariable output error state space (PO-MOESP), and numerical algorithm for subspace state space system identification (N4SID) in their covariance form [28]. In this paper, we will focus on the unweighted IV related to $\mathcal{H}_p$ defined as

$$\mathcal{H}_p = \left\langle \mathcal{Y}^+_{0,p}, \mathcal{L}^-_{0,M} \right\rangle_N \tag{47}$$

where $\mathcal{L}^-_{0,M}$ is obtained, with the notations of Section III-A2, by stacking, for $i = -M, \ldots, 0$

$$L_i \triangleq \mathrm{E}_N \left(W_i \,\middle|\, \left(\mathcal{U}^+_{0,M} \right)^{\perp} \right), \qquad \text{where } W_i \triangleq \begin{bmatrix} U_i \\ Y_i \end{bmatrix}. \tag{48}$$

In our general setting, this amounts to starting from matrices

$$R_i = \langle Y_i, \mathcal{L}^-_{0,M} \rangle_N. \tag{49}$$

The parallel with algorithm $R^{(42)}$ can be stated as follows: L_i in (48) corresponds to Y_i in (41). The associated instrument z_k is, therefore, the sequence of the successive columns of matrix $\mathcal{L}^-_{0,M}$. Note that z_k is $\mathcal{F}^o_k$-measurable, which proves point (27) of Condition 2. The rest of the analysis is the same as for $R^{(42)}$.

We can parallel the variant $R^{(43)}$ of $R^{(42)}$; just take

$$Z_0(N) \triangleq \mathrm{E}_N \left(Y_0 \,\middle|\, \left(\mathcal{U}^+_{0,M} \right)^{\perp} \right)$$

and

$$\hat{r}_i(N) = \langle Y_i, Z_0 \rangle_N.$$

Thus

$$R_i(N) = [\, \hat{r}_i(N) \quad \cdots \quad \hat{r}_{i+M}(N) \,]. \tag{50}$$

Algorithm $R^{(50)}$ relates to $R^{(49)}$ in the very same way that $R^{(43)}$ relates to $R^{(42)}$.

b) Covariance-driven IO subspace algorithms with projection on the orthogonal of the input [14]: This algorithm consists in computing, for $i = 1, \ldots, p$ [cf. notations (10)]

$$\begin{aligned} \hat{r}_i(N) &= \langle Y_i, Z_0 \rangle_N, \qquad \text{where } Z_0(N) \triangleq \mathrm{E}_N\left(Y_0 \,|\, \mathcal{U}_{0,M}^{\perp}\right) \\ R_i(N) &= [\hat{r}_i(N) \quad \hat{r}_{i+1}(N) \quad \cdots \quad \hat{r}_{i+M}(N)]. \end{aligned} \tag{51}$$

This algorithm relates to $R^{(50)}$ by simply changing the instrument. Only point (27) of Condition 2 regarding measurability needs to be checked, which is immediate.

c) Data-driven IO subspace algorithm with projection on the orthogonal of the input [23]: This algorithm is known as the projection algorithm in [23, Ch. 2.3.2.]. It consists in computing $\mathcal{H}_p = \mathrm{E}_N(\mathcal{Y}_{0,p}^{+} \,|\, \mathcal{Z}_{0,M}^{-})$, where Z_i is as in (51). We conclude as for $\mathcal{H}_p^{(45)}(N)$.

d) Data-driven subspace algorithm using projected inputs as instruments [10], [25], [26]: This algorithm was first proposed in [25] and [26] under the name of past input multivariable out put error state space (PI-MOESP). It was studied recently in [10] and [11]; a detailed presentation is found in [11]. It consists in computing a (left) weighted version of

$$\mathcal{H}_p = \mathrm{E}_N\left(\mathcal{Y}_{0,p}^{+} \,|\, \mathcal{L}_{0,M}^{-}\right) \tag{52}$$

where $\mathcal{L}_{0,M}^{-}$ is obtained, with the notations of Section III-A2, by stacking

$$L_i \triangleq \mathrm{E}_N\left(U_i \,\middle|\, \left(\mathcal{U}_{0,M}^{+}\right)^{\perp}\right), \qquad \text{for } i = -M, \ldots, 0. \tag{53}$$

Introduce the following instrument:

$$Z_0 = (\langle \mathcal{L}_{0,M}^{-}, \mathcal{L}_{0,M}^{-} \rangle_N)^{-\frac{1}{2}} \mathcal{L}_{0,M}^{-}(N). \tag{54}$$

The squaring argument already used in analyzing (45) can be used here as well. Once more, instrument (54) satisfies the measurability property (27) in Condition 2, and we conclude as for $\mathcal{H}_p^{(45)}(N)$. Note that, to get this measurability condition, it was essential that the observation σ-algebra $\mathcal{F}_k^o$ contains both past and future of the observed input u.

Remark: Note that the same analysis would work if the following substitution was made in (53):

$$U_i \text{ is replaced by } \begin{bmatrix} U_i \\ Y_i \end{bmatrix} \tag{55}$$

while keeping $(\mathcal{U}_{0,M}^{+})^{\perp}$ unchanged.

e) Data-driven subspace identification using oblique projections [22], [23]: This category includes popular subspace algorithms such as N4SID and MOESP [22], [23] as well as any variation of them by using weights, including the CVA method [23]. We focus on N4SID and MOESP.

The popular N4SID algorithm of [22] and [23, Sec. 4.3.1], consists in computing the so-called oblique projection of $\mathcal{Y}_{0,p}^{+}$ on $\begin{bmatrix} \mathcal{U}_{0,M}^{-} \\ \mathcal{Y}_{0,M}^{-} \end{bmatrix}$ along $\mathcal{U}_{0,M}^{+}$

$$\mathcal{H}_p(N) = \mathcal{Y}_{0,p}^{+} \Big/_{\mathcal{U}_{0,M}^{+}} \begin{bmatrix} \mathcal{U}_{0,M}^{-} \\ \mathcal{Y}_{0,M}^{-} \end{bmatrix}. \tag{56}$$

Formula (56) rewrites as follows [23, Sec. 4.3.1]:

$$\mathcal{H}_p(N) = \mathcal{H}_p^{w}(N) \langle \mathcal{L}_{0,M}^{-}, \mathcal{L}_{0,M}^{-} \rangle_N^{\dagger} \mathcal{W}_{0,M}^{-} \tag{57}$$

where

$$\mathcal{H}_p^{w}(N) \triangleq \langle \mathcal{Y}_{0,p}^{+}, \mathcal{L}_{0,M}^{-} \rangle_N \tag{58}$$

and $\mathcal{W}_{0,M}^{-}$ and $\mathcal{L}_{0,M}^{-}$ are obtained, with the notations of Section III.A.2, by stacking, for $i = -M, \ldots, 0$, the vectors W_i and L_i defined in (48). Introduce [compare with (57)]

$$\mathcal{H}_p'(N) \triangleq \mathcal{H}_p^{w}(N) \langle \mathcal{L}_{0,M}^{-}, \mathcal{L}_{0,M}^{-} \rangle_N^{\dagger} \langle \mathcal{W}_{0,M}^{-}, \mathcal{W}_{0,M}^{-} \rangle_N^{1/2}. \tag{59}$$

Set $\mathcal{K}(N) = \mathcal{H}_p(N)\mathcal{H}_p(N)^T$ and $\mathcal{K}'(N) = \mathcal{H}_p'(N)\mathcal{H}_p'(N)^T$. Note that $\mathcal{K}(N) = \mathcal{K}'(N)$. By using repeatedly the squaring argument of Corollary 1, we deduce that, if $\mathcal{H}_p'(N)$ satisfies Condition 1, then $\mathcal{H}_p(N)$ provides consistent estimators for (C, A). Next, notice that $\mathcal{H}_p^{w}(N)$ in (58) coincides with (47), so we already know that it satisfies Condition 1. Finally, according to Lemma 2, the pair of matrices $(C(N), A(N))$ corresponding to $\mathcal{H}_p'(N)$ will be consistent provided that (23) holds, with $W_\lambda(N) = I$.

The following remark can be stated about (30) and (23). These conditions are fragile if the postweighting in (59) is close to having rank less than p, which happens when the future $\mathcal{U}_{0,M}^{+}$ of input u is almost parallel to $\mathcal{W}_{0,M}^{-}$. The latter fact is indeed known from the practice about N4SID and is also analyzed in [12].

The same analysis also applies to the MOESP algorithm described in [23, Sec. 4.3.2], [25], [26], and [3]. This algorithm consists in computing

$$R_i(N) = R_i^{(56)}(N) \Pi_{(\mathcal{U}_{0,M}^{+})^{\perp}}$$

where $\Pi_{(\mathcal{U}_{0,M}^{+})^{\perp}}$ denotes the (orthogonal) projection on $(\mathcal{U}_{0,M}^{+})^{\perp}$. Thus, MOESP amounts to computing

$$\mathcal{H}_p(N) = \mathcal{H}_p^{w}(N) \langle \mathcal{L}_{0,M}^{-}, \mathcal{L}_{0,M}^{-} \rangle_N^{\dagger} \mathcal{L}_{0,M}^{-}. \tag{60}$$

Following the same lines as for N4SID, MOESP yields consistent estimates. The same remark as for N4SID applies, regarding the conditioning of the postweighting and its impact on the behavior of the algorithm. The class of subspace methods described in [23], including CVA, is analyzed along the same lines.

C. Time- Versus Frequency Domain

For $(y_k)_{k \in \mathbb{Z}}$, an $\mathbb{R}^q$-valued data sequence and $N > 0$, a window length, the discrete Fourier transform (DFT) of $Y_i(N)$, denoted by $\hat{Y}_i(N)$, is equal to

$$\hat{Y}_i(N) = Y_i(N) \Delta_N^q \tag{61}$$

where [in (62), $\otimes$ denotes the Kronecker product]

$$\Delta_N^q \triangleq \frac{1}{N^{q/2}} \begin{bmatrix} e^{-2j\pi \frac{0}{N}} & \cdots & e^{-2j\pi \frac{0N}{N}} \\ e^{-2j\pi \frac{1}{N}} & \cdots & e^{-2j\pi \frac{1N}{N}} \\ \vdots & \vdots & \vdots \\ e^{-2j\pi \frac{N-1}{N}} & \cdots & e^{-2j\pi \frac{(N-1)N}{N}} \end{bmatrix} \otimes I_q. \tag{62}$$

Since matrix Δ_N^q is orthogonal, then

$$\langle \hat{X}, \hat{Y} \rangle_N = \langle X, Y \rangle_N$$

and

$$\mathrm{E}_N(\hat{X} \,|\, \hat{Y}) = \mathrm{E}_N(X \,|\, Y)\Delta_N^q.$$

Hence, Condition 2 can be considered equivalently in the time domain or in the frequency domain. Therefore, frequency-domain subspace algorithms corresponding to [17] and [18] behave exactly the same way as their time-domain counterparts regarding nonstationary consistency.

V. Conclusion

In this paper, we have revisited eigenstructure identification via subspace methods. This problem is clearly easier than full system matrix identification. On the other hand, consistency of eigenstructure identification still holds for nonstationary inputs (in fact, for "nonstationary zero part").

For this paper, we have adapted the original method of [6]. We believe that our presentation enlightens the reasons for subspace methods to converge, and therefore, can serve as a guideline for further new designs. Our analysis shows that the old fashioned "instruments" are still a useful concept in this respect.

Martingale techniques were used to deal with unobserved inputs—for unobserved inputs, "deterministic" projections based on observed data cannot be used; they can be replaced by "stochastic" projections via conditional expectations. This technique requires a probabilistic setting for the unobserved inputs, and the white noise assumption provides a situation in which finding instruments is easy. This suggests that our martingale approach could possibly be replaced by any other method providing orthogonality conditions without the need for observing data.

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.

Finally, we have only studied nonstationary consistency, not nonstationary convergence rates. The latter subject is definitely much harder. The only results we are aware of in this direction are found in [19].

Appendix
Missing proofs of Section III-A4

A. Proof of Lemma 2

The proof relies on the Lemma 2 in [6, App. C], which we repeat here for completeness.

Lemma 4 [6]: Let $\mathcal{K}(N)$ and $\mathcal{K}'(N)$ be two sequences of matrices of fixed dimensions, satisfying the following conditions:

1) sequence $\mathcal{K}(N)$ is bounded and $\mathcal{K}(N) - \mathcal{K}'(N) \to 0$ when $N \to \infty$;
2) for every N, the SVD of matrix $\mathcal{K}(N)$ is $\mathcal{K}(N) = \mathbf{U}\,\mathrm{diag}(\sigma_1, \ldots, \sigma_n)\,\mathbf{V}^T$, and $\liminf_{N\to\infty} \sigma_n > 0$ holds.

By an SVD-decomposition, rewrite $\mathcal{K}'(N)$ as $\mathcal{K}'(N) = \mathbf{U}'\,\mathrm{diag}(\sigma'_1, \ldots, \sigma'_n)\,\mathbf{V'}^T +$ higher order terms. Then, there exists a sequence of $p \times p$-matrices $T(N)$, bounded with bounded inverse, such that $\mathbf{U}'(N) - \mathbf{U}(N)T(N) \to 0$ when $N \to \infty$.

Return to the proof of Lemma 2. Set $\mathcal{K}'(N) = \mathcal{H}_p(N)$ and

$$\mathcal{K}(N) = W_\lambda(N)\mathbf{U}_w \mathrm{diag}\,(\sigma_1^w, \ldots, \sigma_n^w)\,\mathbf{V}_w^T W_\rho^T(N).$$

By (22) and (23), we have $\mathcal{K}(N) - \mathcal{K}'(N) \to 0$. On the other hand, it is assumed for Lemma 2 that $\mathcal{K}'(N) = \mathcal{H}_p(N)$ is bounded. Therefore, Lemma 4 applies. Since $\mathcal{K}(N)$ has rank exactly p, the left most factor in the SVD of $\mathcal{K}(N)$ is obtained from $W_\lambda(N)\mathbf{U}_w$ by a postmultiplication by an invertible matrix. On the other hand, the left factor $\mathbf{U}'$ associated to $\mathcal{K}'(N) = \mathcal{H}_p(N)$ by Lemma 4 coincides with $\mathbf{U}$ in formula (20). Hence, $W_\lambda(N)\mathbf{U}_w$ and $\mathbf{U}$ are related via the postmultiplication by an invertible matrix. From this, the conclusion of Lemma 2 follows.

B. Proof of Corollary 1

For A, a matrix, and n, an integer, denote by $[A]_{\le n}$ the matrix obtained by zeroing all singular values of rank $>n$ in the SVD of A, and set $[A]_{>n} = A - [A]_{\le n}$. We successively prove points 1 and 2.

Consider first point 1. Since $\mathcal{H}_p(N)$ satisfies Condition 1, then

$$\left[\mathcal{H}_p^w(N)\mathcal{H}_p^w(N)^T\right]_{>n} = [\mathcal{H}_p(N)]_{>n} \to 0 \tag{63}$$

holds. By the orthogonality property of the SVD, we have

$$\left[\mathcal{H}_p^w(N)\right]_{>n} \mathcal{H}_p^w(N)^T = \left[\mathcal{H}_p^w(N)\mathcal{H}_p^w(N)^T\right]_{>n}$$

whence

$$\left[\mathcal{H}_p^w(N)\right]_{>n} \mathcal{H}_p^w(N)^T \to 0. \tag{64}$$

Matrices $\mathcal{H}_p(N)$ and $\mathcal{H}_p^w(N)$ are related as in Lemma 2 with $W_\lambda(N) = I$ and $W_\rho(N) = \mathcal{H}_p^w(N)$. With this choice for the weights, (64) is exactly (23). On the other hand, since $\mathcal{H}_p(N)$ satisfies Condition 1, then, by Theorem 3, the pair $(C(N), A(N))$ is consistent. Thus, Lemma 2 applies and yields the consistency of $(C_w(N), A_w(N))$.

Consider now point 2. Since $\mathcal{H}_p^w(N)$ satisfies Condition 1, it follows that $[\mathcal{H}_p^w(N)]_{>n} \to 0$, which implies (63), and thus also (64). Since $\mathcal{H}_p^w(N)$ satisfies Condition 1, then, by Theorem 3, the pair $(C^w(N), A^w(N))$ is consistent; and we conclude again, by a reverse use of Lemma 2.

Acknowledgment

The authors would like to thank M. Basseville for corrections and useful suggestions on an earlier version of this paper. They would also like to thank one of the reviewers who significantly helped in improving the final version of this paper.

References

[1] M. Basseville, A. Benveniste, M. Goursat, L. Hermans, L. Mevel, and H. Van der Auweraer, "Output-only subspace-based structural identification: From theory to industrial testing practice," *ASME J. Dyn. Syst. Meas. Control*, vol. 123, no. 4, pp. 668–676, Dec. 2001.

[2] D. Bauer, M. Deistler, and W. Scherrer, "On the impact of weighting matrices in subspace algorithms," in *Proc. IFAC Int. Symp. Identif.*, Santa Barbara, CA, 2000, Paper no. 16.
[3] D. Bauer and M. M. Jansson, "Analysis of the asymptotic properties of the MOESP type of subspace algorithms," *Automatica*, vol. 36, pp. 497–509, 2000.
[4] D. Bauer, M. Deistler, and W. Scherrer, "Consistency and asymptotic normality of some subspace algorithms without observed inputs," *Automatica*, vol. 35, pp. 1243–1254, 2002.
[5] D. Bauer and L. Ljung, "Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms," *Automatica*, vol. 36, pp. 763–773, 2002.
[6] A. Benveniste and J.-J. Fuchs, "Single sample modal identification of a nonstationary stochastic process," *IEEE Trans. Autom. Control*, vol. AC-30, no. 1, pp. 66–74, Jan. 1985.
[7] B. Cauberghe, P. Guillaume, P. Verboven, and E. Parloo, "Identification of modal parameters including unmeasured forces and transient effects," *J. Sound Vib.*, vol. 265-3, pp. 609–625, Aug. 2003.
[8] B. Cauberghe, "Applied frequency domain system identification in the field of experimental and operational modal analysis," Ph.D. dissertation, Dept. Mech. Eng., Vrije Universitet Brussel, Brussel, Belgium, 2004.
[9] B. Cauberghe, P. Guillaume, and R. Pintelon, "Combined deterministic-stochastic frequency domain subspace," *Automatica*, submitted for publication.
[10] A. Chiuso and G. Picci, "Subspace identification by data orthogonalization and model decoupling," *Automatica*, vol. 40, pp. 1689–1704, 2004.
[11] A. Chiuso and G. Picci, "Asymptotic variance of Subspace methods by data orthogonalization and model decoupling: A comparative analysis," *Automatica*, vol. 40, pp. 1705–1717, 2004.
[12] A. Chiuso and G. Picci, "On the ill conditioning of subspace identification with inputs," *Automatica*, vol. 40, pp. 575–589, 2004.
[13] M. Deistler, K. Peternell, and W. Scherrer, "Consistency and relative efficiency of subspace methods," *Automatica*, vol. 31, pp. 1865–1875, 1995.
[14] A. Guyader and L. Mevel, "Covariance driven subspace methods: Input/output vs. output-only," in *Proc. Int. Modal Anal. Conf.*, Orlando, FL, Feb. 2003, Paper no. 136.
[15] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Applications*. New York: Academic, 1980.
[16] W. E. Larimore, "System identification, reduced order filters and modelling via canonical variate analysis," in *Proc. Amer. Control Conf.*, San Francisco, CA, Jun. 1983, pp. 445–451.
[17] T. McKelvey, "Frequency domain system identification with instrumental based algorithm," in *Proc. ASME Design Eng. Tech. Conf. (DETC)*, Sacramento, CA, Sep. 1997, p. VIB-4252.
[18] T. McKelvey, H. Akcay, and L. Ljung, "Subspace based multivariable system identification from frequency response data," *IEEE Trans. Autom. Control*, vol. 41, no. 7, pp. 960–978, Jul. 1996.
[19] G. Moustakides and A. Benveniste, "Detecting changes in the AR part of a nonstationary ARMA process," *Stochastics*, vol. 16, pp. 137–155, 1986.
[20] J. Neveu, *Discrete-Parameter Martingales*. New York: Elsevier, 1975.
[21] M. Prevosto, M. Olagnon, A. Benveniste, M. Basseville, and G. Le Vey, "State-space formulation, a solution to modal parameter estimation," *J. Sound Vib.*, vol. 148, no. 2, pp. 329–342, 1991.
[22] P. Van Overschee and B. DeMoor, "N4Sid: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, pp. 75–93, 1994.
[23] P. Van Overschee and B. DeMoor, *Subspace Identification for Linear Systems*. Norwell, MA: Kluwer, 1996.
[24] V. Verdult and M. Verhaegen, "Subspace Identification of multivariable linear parameter-varying systems," *Automatica*, vol. 38, pp. 805–814, 2002.
[25] M. Verhaegen and P. Dewilde, "Subspace model identification part 3. Analysis of the ordinary output-error state space model identification algorithm," *Int. J. Control*, vol. 55, pp. 555–586, 1993.
[26] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovation forms from input-output data," *Automatica*, vol. 30, pp. 61–74, 1994.
[27] M. Viberg, "Subspace-based methods for the identification of linear time-invariant systems," *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995.
[28] M. Viberg, B. Wahlberg, and B. Ottersen, "Analysis of state space system identification methods based on instrumental variables and subspace fitting," *Automatica*, vol. 33, pp. 1603–1616, 1997.

Albert Benveniste (F'91) received the These d'Etat in mathematics, probability theory, from Ecole des Mines de Paris, Paris, France, in 1975.

From 1976 to 1979, he was an Associate Professor in mathematics at the Universite de Rennes I, Rennes, France. Since 1979, he has been the Directeur de Recherche at INRIA-Rennes, France. He has coauthored with M. Metivier and P. Priouret the book *Adaptive Algorithms and Stochastic Approximations* (New York: Springer-Verlag, 1990). He has published over 100 journal papers and several hundreds conference papers. He is the coinventor, with P. Le Guernic, of the synchronous language Signal for embedded systems design. Since 1996, he has been active in distributed algorithms for network and service management in telecommunications, where he has contributed to distributed fault diagnosis. His current interests include system identification and change detection in signal processing and automatic control, vibration mechanics, reactive and real-time embedded systems design in computer science, and network and service management in telecommunications.

Dr. Benveniste is or has been an Associate Editor for several journals and the Program Committee (PC) member of many conferences in computer science. In 1980, he received the IEEE TRANSACTIONS ON AUTOMATIC CONTROL Best Paper Award for his paper on blind deconvolution in data communications. In 1990, he received the CNRS silver medal.

Laurent Mevel received the Ph.D. degree in applied mathematics in the field of statistical inference for hidden Markov models from the University of Rennes, Rennes, France, in 1997.

Currently, he is with IRISA, Rennes, France, where he is the Chargé de Recherches at the French National Institute for Research in Computer Science and Automatic Control (INRIA). He is working on identification and detection of partially hidden stochastic systems. His interests include structural health monitoring and detection for vibration mechanics, modeling and rejection of environmental effects, model validation and uncertainty estimations, and model reduction of large systems.