ELSEVIER

Signal Processing 52 (1996) 145–160

SIGNAL PROCESSING

Stochastic realization with exogenous inputs and 'subspace-methods' identification

Giorgio Picci,[a,*,1] Tohru Katayama[b]

[a] *Dipartimento di Elettronica e Informatica and LADSEB-CNR, University of Padova, 35131 Padua, Italy*
[b] *Department of Applied Mathematics and Physics, Kyoto University, Kyoto 606-01, Japan*

Received 16 June 1995; revised 27 February 1996

Abstract

In this paper we study stochastic realization of stationary processes with exogenous inputs in the absence of feedback and we briefly discuss its application to identification. In particular, we derive and characterize the family of minimal state-space models of such processes and introduce a very natural block structure which is generically minimal. This model structure leads very naturally to 'subspace'-based identification algorithms which have a simpler structure of those existing in the literature.

Zusammenfassung

Dieser Artikel untersucht stochastische Realisierungen von stationären Prozessen mit exogenen Eingängen in Abwesenheit von Rückkopplungen. Weiters werden Anwendungen auf die Identifikation kurz diskutiert. Im besonderen wird die Familie minimaler Zustandsraummodelle solcher Prozesse abgeleitet und charakterisiert, und es wird eine sehr natürliche und minimale Blockstruktur eingeführt. Diese Modellstruktur führt auf natürliche Weise zu auf Unterräumen basierenden Identifikationsalgorithmen, die eine einfachere Struktur als die in der Literatur vorhandenen Algorithmen aufweisen.

Résumé

Dans cet article nous étudions la réalisation stochastique de processus stationnaires avec des entrées exogènes dans des conditions d'absence de retour vers l'entrée et nous discutons brièvement ses applications au problème de l'identification. En particulier, nous dérivons et caractérisons la famille des modèles d'espace d'état minimum pour ce type de processus, et introduisons une structure par blocs très naturelle, qui est génériquement minimale. Cette structure de modèles conduit très naturellement à des algorithmes d'identification basés sur les "sous-espaces", qui ont une structure plus simple que ceux existant dans la littérature.

Keywords: Stochastic realization; Stochastic systems with exogenous inputs; Subspace-methods identification; Canonical correlation analysis

* Corresponding author. E-mail: picci@dei.unipd.it.

[1] Part of this work was done while the first author was supported by a Kyoto University Foundation Fellowship in the Fall 1994.

PII S0165-1684(96)00050-3

1. Introduction

A number of recent papers [12, 19, 29–32] concerns identification of state-space models by geometric methods based on ideas from stochastic realization theory. In these so-called 'subspace methods', the state space of the model is first constructed by geometric operations on the vector spaces generated by the observed data (projection of the future onto the past, etc.). Singular-value decomposition techniques are then applied to find well-conditioned bases in the state space which generically correspond to balanced canonical forms for the state-space model to be identified. Using the state estimates constructed in this way the system matrices can then be computed directly in balanced canonical form thus avoiding the difficult structure-selection problem inherent in the classical approaches to multivariable identification.

Now, while subspace-methods identification of 'purely stochastic' systems (i.e. of signals or time series) seems to be reasonably well-understood, see e.g. the book [3], the influential paper [29] and the subsequent discussions in [16], for signals which are driven by 'inputs' or, better, *exogenous* variables, the picture looks still a bit unsatisfactory, as various algorithms are given in the literature [20, 32, 30], some of which require a rather complicated analysis to motivate [30]. The assumptions on the input signal are different (sometimes assumed to be a white-noise process, a 'deterministic' known signal, etc.) in particular the sharp distinction made in the literature between 'deterministic' inputs and other 'stochastic' signals in the model may be puzzling for the user. How is he going to decide in practice if an input signal is 'stochastic' or 'deterministic'?

What seems to be needed is an abstract geometric procedure for solving the stochastic realization problem *with inputs* in a probabilistic setup (similar to what is available in the literature for processes or time series, see e.g. [13, 14, 16, 17]). This is a conceptual prerequisite to subspace-based identification methods and it seems to us that only a better understanding of realization theory would eventually make possible the understanding of and a comparison between the different methods and algorithms existing in the literature.

Modeling of processes influenced by exogenous inputs, or in more precise terms, geometric realization theory of processes with exogenous inputs, is not discussed in the literature in a satisfactory way. So far only the white-noise input case is treated in [28] but the setting of [28] is very much coordinate-dependent and not geometric in spirit.

The purpose of this paper is to provide some basic ideas for realization of processes with exogenous inputs and to show how this theory translates very naturally into subspace-based identification methods which are simple and easy to understand. The theory and the proposed methods are based on the assumption of *absence of feedback* from y to u which somehow is equivalent to the declaration of 'determinstic' input signals. The motivations and a precise statement of the feedback-free condition are discussed in Sections 2 and 3.

In Sections 4 and 5 we derive by a 'canonical' geometric procedure the family of minimal state-space models of such processes and introduce a very natural block structure which is generically minimal. Using this model structure the identification problem with exogenous inputs can be split into two completely separate subproblems of identification of two, suitably defined, 'stochastic' and 'deterministic' components of the signal y. The algorithms resulting from this model class have therefore a convenient modular structure which is not evident in any of the papers appeared so far.

The stochastic identification subproblem is exactly of the standard 'time-series' type (no inputs) and can of course be attacked by, e.g., the method of [29]. The other 'deterministic' identification problem can be solved by the 'deterministic' methods presented in [20, 32] or suitable variants of them which will be briefly discussed at the end of this paper.

In the last section some hints for a possible numerical implementation are given.

1.1. Background and notations

Let $u := \{u(t)\}$ and $y := \{y(t)\}$ be jointly stationary (wide sense) discrete-time vector processes with zero mean and finite variance of dimensions p and m and let $\mathscr{U}$ and $\mathscr{Y}$ denote the (infinite-dimensional) Hilbert spaces of scalar finite variance random variables obtained first by taking all finite linear combinations $\{\sum a_k' u(t_k) \mid a_k \in \mathbb{R}^p,\ t_k \in \mathbb{Z}\}$ and $\{\sum a_k' y(t_k) \mid a_k \in \mathbb{R}^m,\ t_k \in \mathbb{Z}\}$, respectively, and then closing the two

vector spaces with respect to the norm induced by the scalar product $\langle \xi, \eta \rangle = E\{\xi\eta\}$, where $E\{\cdot\}$ denotes mathematical expectation. The spaces $\mathscr{U}$ and $\mathscr{Y}$ are commonly denoted by $\overline{\text{span}}\{u(t) \mid t \in \mathbb{Z}\}$ and $\overline{\text{span}}\{y(t) \mid t \in \mathbb{Z}\}$. They contain all linear functionals (i.e. all linear scalar statistics) of the 'history' of the processes u and y. The *infinite past* and *future* subspaces at time t are defined as

$$\mathscr{U}_t^- := \overline{\text{span}}\{u(s) \mid s < t\},$$

$$\mathscr{Y}_t^- := \overline{\text{span}}\{y(s) \mid s < t\},$$

and

$$\mathscr{U}_t^+ := \overline{\text{span}}\{u(s) \mid s \geqslant t\},$$

$$\mathscr{Y}_t^+ := \overline{\text{span}}\{y(s) \mid s \geqslant t\},$$

respectively. Note that, according to a widely accepted convention, the present is included in the future only and not in the past.

Normally, all linear operations on the data will be time-invariant. By stationarity we could then fix the present instant of time to an arbitrary value say $t = 0$ and then propagate everything in time by the action of the *shift* operator σ, where $\sigma(\sum a_k' u(t_k) + \sum b_k' y(t_k)) := \sum a_k' u(t_k + 1) + \sum b_k' y(t_k + 1)$. By stationarity, the shift is a norm preserving linear operator which can naturally be extended to the space of all joint linear functionals (statistics) of u and y.

In what follows, the symbols $\vee$, $+$ and $\oplus$ will denote vector sum, *direct* vector sum and *orthogonal* vector sum of subspaces, the symbol $\mathscr{A}^\perp$ will denote the orthogonal complement of the subspace $\mathscr{A}$ with respect to some predefined ambient space, usually $\mathscr{U} \vee \mathscr{Y}$. The orthogonal projection onto the subspace $\mathscr{A}$ will be denoted by the symbol $E(\cdot|\mathscr{A})$ or by the shorthand $E^{\mathscr{A}}$. The notation $E(z|\mathscr{A})$ will be used also when z is vector-valued. The symbol will then just denote the vector with components $E(z_k|\mathscr{A})$, $k = 1, \ldots$

The notation $\mathscr{A} \perp \mathscr{B}|X$ means that the two subspaces $\mathscr{A}$ and $\mathscr{B}$ are *conditionally orthogonal* given a third subspace X, i.e.

$$\langle \alpha - E^X\alpha, \beta - E^X\beta \rangle = 0 \quad \text{for } \alpha \in \mathscr{A},\ \beta \in \mathscr{B}. \tag{1.1}$$

When $X = 0$, this reduces to the usual orthogonality $\mathscr{A} \perp \mathscr{B}$. Conditional orthogonality is orthogonality after subtracting the projections onto X. Since $\alpha = E^X\alpha + E^{X^\perp}\alpha$, (1.1) is actually the same thing as

$$\langle E^{X^\perp}\alpha, \beta - E^X\beta \rangle = \langle E^{X^\perp}\alpha, \beta \rangle = 0 \quad \text{for } \alpha \in \mathscr{A},\ \beta \in \mathscr{B}. \tag{1.2}$$

The following proposition is taken from [13].

Proposition 1.1. *The following statements are equivalent*:

(i) $\mathscr{A} \perp \mathscr{B} \mid X$,
(ii) $\mathscr{B} \perp \mathscr{A} \mid X$,
(iii) $(\mathscr{A} \vee X) \perp \mathscr{B} \mid X$,
(iv) $E^{\mathscr{A} \vee X}\beta = E^X\beta$ *for* $\beta \in \mathscr{B}$,
(v) $(\mathscr{A} \vee X) \ominus X \perp \mathscr{B}$,
(vi) $E^{\mathscr{A}}\beta = E^{\mathscr{A}}E^X\beta$ *for* $\beta \in \mathscr{B}$.
(*Here* $\mathscr{C} \ominus X$ *is the orthogonal complement of* X *in* $\mathscr{C}$).

2. Feedback-free processes

Following Granger [10] and subsequent work by Caines, Chan, Anderson, Gevers etc. [4, 2, 7, 8], we shall say that *there is no feedback from y to u* if the future of u is conditionally uncorrelated (which is the same as independent in the Gaussian case) from the past of y given the past of u itself. In our Hilbert space setup this is written as

$$\mathscr{U}_t^+ \perp \mathscr{Y}_t^- \mid \mathscr{U}_t^-. \tag{2.1}$$

From condition (iii) in Proposition 1.1, written for $\mathscr{A} = \mathscr{U}_t^+$, the feedback-free condition is seen to be equivalent to $\mathscr{Y}_t^- \perp \mathscr{U} | \mathscr{U}_t^-$ and hence, from (iv), to $E[\mathscr{Y}_t^- | \mathscr{U}] = E[\mathscr{Y}_t^- | \mathscr{U}_t^-]$, so that

$$E[y(t)|\mathscr{U}] = E[y(t)|\mathscr{U}_{t+1}^-] \quad \text{for all } t \in \mathbb{Z}, \tag{2.2}$$

which yields the well-known equivalence between absence of feedback (from y to u) and *causality* of the estimator $E[y(t)|\mathscr{U}]$. This equivalence is discussed for example in [25].

It also follows from the equality between causal and non-causal estimates that

$$y_s(t) := y(t) - E[y(t)|\mathscr{U}_{t+1}^-] = y(t) - E[y(t)|\mathscr{U}] = E[y(t)|\mathscr{U}^\perp], \tag{2.3}$$

so that $y_s(t) \perp \mathscr{U}$ for all t, i.e. the 'causal estimation error' is uncorrelated with the whole history of the input process u. We shall call the process y_s the *stochastic component of* y (this corresponds to the 'stochastic component' y^s of [30]). Similarly, the stochastic process y_u defined by the complementary projection

$$y_u(t) := E[y(t)|\mathscr{U}], \quad t \in \mathbb{Z},$$

is named the *deterministic component of* y.

To avoid confusion with time indexing in the notations we shall denote by $\tilde{\mathscr{Y}}$ the Hilbert subspace of $\mathscr{U} \vee \mathscr{Y}$ linearly generated by $\{y_s(t) \mid t \in \mathbb{Z}\}$ and by $\hat{\mathscr{Y}}$ the Hilbert subspace of $\mathscr{U}$ linearly generated by $\{y_u(t) \mid t \in \mathbb{Z}\}$. Note that not only we have $\mathscr{Y} \vee \mathscr{U} = \tilde{\mathscr{Y}} \oplus \mathscr{U}$, but in virtue of causality,

$$\mathscr{Y}_t^- \vee \mathscr{U} = \tilde{\mathscr{Y}}_t^- \oplus \mathscr{U}$$

for all t, where $\tilde{\mathscr{Y}}_t^-$ is the past space of the process y_s at time t. We stress that the stochastic and 'deterministic' components in the decomposition

$$y(t) = y_s(t) + y_u(t) \tag{2.4}$$

are *completely uncorrelated*, i.e. $E[y_s(t)y_u(\tau)'] = 0$ for all $t, \tau \in \mathbb{Z}$.

If there is no causality (or equivalently if there is feedback from y to u) the very notion of 'input' loses its meaning since, as shown, e.g., in Ref. [7] the variable $u(t)$ is then also determined by a dynamical relation involving its own past and the 'output' process y which is now playing the role of an exogenous variable to determine u. Correspondingly, as it has been argued in several places in the literature, identification in the presence of feedback (and of course in the absence of any other specific information on the feedback loop) is essentially equivalent to identification of the *joint process* $[y', u']$, in the sense of time-series identification.

3. Stochastic realization

State-space modeling of the joint stationary process (y, u) is a well-understood problem which has been treated in several places in the literature. The *geometric* approach is based on the idea of *Markovian Splitting Subspaces*. Any such subspace $\mathscr{X}_t$ makes the joint past and future spaces $\mathscr{Y}_t^- \vee \mathscr{U}_t^- \vee \boldsymbol{X}_t^-$ and $\mathscr{Y}_t^+ \vee \mathscr{U}_t^+ \vee \boldsymbol{X}_t^+$ conditionally uncorrelated given the present $\boldsymbol{X}_t$ at time t, i.e.

$$\mathscr{Y}_t^- \vee \mathscr{U}_t^- \vee \mathscr{X}_t^- \perp \mathscr{Y}_t^+ \vee \mathscr{U}_t^+ \vee \boldsymbol{X}_t^+ | \mathscr{X}_t. \tag{3.1}$$

This concept embodies the coordinate-free idea of a stochastic *state space* of the joint process in the sense that any state equation representation of the joint process corresponds to a Markovian splitting subspace and conversely, see e.g. [14].

To get models useful in identification, one should moreover look for state spaces which are minimal (i.e. of minimal dimension) and can be constructed from (in fact are contained in) the data space $\mathscr{Y} \vee \mathscr{U}$. An important example of such a minimal Markovian Splitting subspace is the *forward predictor space*

$$\boldsymbol{X}_t^{+/-} := E[\mathscr{Y}_t^+ \vee \mathscr{U}_t^+ | \mathscr{Y}_t^- \vee \mathscr{U}_t^-]. \tag{3.2}$$

This state space leads to state equations of the Kalman-filter type for (y, u), the well-known (forward) innovations representation. From the conditional orthogonality (2.1) we get the following decomposition of the joint predictor space of (y, u). The justification is easy and will be omitted.

Lemma 3.1. *For a feedback-free process, the joint predictor space* (3.2) *is equal to the vector sum*

$$\boldsymbol{X}_t^{+/-} := E[\mathscr{Y}_t^+ | \mathscr{Y}_t^- \vee \mathscr{U}_t^-] \vee E[\mathscr{U}_t^+ | \mathscr{U}_t^-]. \tag{3.3}$$

Note that $E[\mathscr{U}_t^+ | \mathscr{U}_t^-]$ is the (forward) predictor space for the input process alone. This state space carries information only about the dynamics of the process u, which we are not interested in. In identification, one is actually interested in obtaining only dynamic models of the 'output' process y expressing $y(t)$ as a causal functional of past input and output values.

Thus, we need to study realization of the output y with respect to the joint information flow $\mathscr{Y}_t^- \vee \mathscr{U}_t^-$. This in turn means that we need to understand the structure of the state spaces (splitting subspaces) which make the future of the process $\mathscr{Y}_t^+$ conditionally orthogonal to the available past information $\mathscr{Y}_t^- \vee \mathscr{U}_t^-$, see [22, 17]. A natural candidate for this role is again the predictor space [2] $X_t^{+/-} := E[\mathscr{Y}_t^+ | \mathscr{Y}_t^- \vee \mathscr{U}_t^-]$ which

[2] Note the different notation than in (3.2).

appears in the vector-sum (3.3). This is generated by the projections

$$E[y(t+k)\,|\,\mathscr{Y}_t^- \vee \mathscr{U}_t^-] = E[y(t+k)\,|\,\tilde{\mathscr{Y}}_t^- \oplus \mathscr{U}_t^-]$$

$$= E[y(t+k)\,|\,\tilde{\mathscr{Y}}_t^-] + E[y(t+k)\,|\,\mathscr{U}_t^-] \qquad (3.4)$$

for all $k \geqslant 0$. We define now the predictor spaces $\tilde{X}_t^{+/-}$ and $\hat{X}_t^{+/-}$ for the stochastic and deterministic components y_s, y_u

$$\begin{aligned} \tilde{X}_t^{+/-} &:= E[\tilde{\mathscr{Y}}_t^+ | \tilde{\mathscr{Y}}_t^-], \\ \hat{X}_t^{+/-} &:= E[\hat{\mathscr{Y}}_t^+ | \mathscr{U}_t^-], \end{aligned} \qquad (3.5)$$

where the symbols $\tilde{\mathscr{Y}}_t$ and $\hat{\mathscr{Y}}_t$ refer to the stochastic and deterministic components of the process y.

From the decomposition (3.4) we immediately have the following result.

Lemma 3.2. *For feedback-free processes the predictor space $X_t^{+/-}$ satisfies the inclusion*

$$X_t^{+/-} \subset \tilde{X}_t^{+/-} \oplus \hat{X}_t^{+/-}, \qquad (3.6)$$

where the orthogonal direct sum in the second member of (3.6) *is a (in general non-minimal) Markovian splitting subspace for $\mathscr{Y}_t^+$ and $\mathscr{Y}_t^- \vee \mathscr{U}_t^-$.*

The inclusion in (3.6) cannot in general be replaced by equality. For it can be checked in simple examples that for arbitrary subspaces $\mathscr{A}, \mathscr{B}, \mathscr{C}$ the projection $E[\mathscr{A}|\mathscr{B} \oplus \mathscr{C}]$ is not the same as $E[\mathscr{A}|\mathscr{B}] \oplus E[\mathscr{A}|\mathscr{C}]$.

We shall not give here a formal proof of the last statement in the lemma since it will be evident later on that there are Markovian state-space models of y with state space the direct sum on the right-hand member of (3.6). This direct sum may however correspond to a redundant description of the process for other reasons which will be clarified in the next section.

Recall that a well-known necessary and sufficient condition for y_s to admit a finite-dimensional state-space model is that the relative predictor space $\tilde{X}_t^{+/-}$ be finite-dimensional. In fact $\tilde{n} := \dim[\tilde{X}_t^{+/-}]$ will then be the dimension of any *minimal* state-space model of y_s. The particular realization having as state space the predictor space $\tilde{X}_t^{+/-}$ will have the form

$$x_s(t+1) = A_s x_s(t) + B_s e_s(t), \qquad (3.7a)$$

$$y_s(t) = C_s x_s(t) + e_s(t), \qquad (3.7b)$$

where $e_s(t)$ is the one-step prediction error of the process y_s based on its own past $\tilde{\mathscr{Y}}_t^-$, i.e. the (forward) innovation process of y_s.

Proposition 3.1. *The innovation of the process y_s is the conditional innovation of y given observations of u up to the present time, i.e.,*

$$\begin{aligned} e_s(t) &= y(t) - E[y(t)\,|\,u(s), y(s-1);\ s \leqslant t] \\ &= y(t) - E[y(t)\,|\,\mathscr{U}_{t+1}^- \vee \mathscr{Y}_t^-]. \end{aligned} \qquad (3.8)$$

Proof. Follows from the chain of equalities,

$$\begin{aligned} &y(t) - E[y(t)\,|\,\mathscr{U}_{t+1}^- \vee \mathscr{Y}_t^-] \\ &\quad = y(t) - E[y(t)\,|\,\mathscr{U}_{t+1}^- \oplus \tilde{\mathscr{Y}}_t^-] \\ &\quad = (y(t) - E[y(t)\,|\,\mathscr{U}_{t+1}^-) - E[y(t)\,|\,\tilde{\mathscr{Y}}_t^-] \\ &\quad = y_s(t) - E[y_s(t) + y_u(t)\,|\,\tilde{\mathscr{Y}}_t^-] \\ &\quad = y_s(t) - E[y_s(t)\,|\,\tilde{\mathscr{Y}}_t^-]. \qquad \square \end{aligned}$$

This observation essentially settles the question of modeling the stochastic component.

4. Realization of the deterministic component

State-space modeling of the deterministic component y_u may be based on the same principle of constructing splitting subspaces for the future of the process y_u and the past of the input process u as discussed in the previous sections. However this approach leads to state-space models which are driven by white noise and 'include' also the dynamics of the process u, which is not interesting for identification and we do not want to appear explicitly in the model. For example, it was shown in [22] that one may choose as state space for y_u the predictor space $E[\hat{\mathscr{Y}}_t^+|\mathscr{U}_t^-]$, but that this choice leads to an innovation model for y_u where the state process is driven by the forward innovation process of u. This model in general includes as a cascade subsystem a state-space innovation representation for u.

In order to construct (non-Markovian) state-space descriptions of y_u driven directly by u and therefore not involving the particular dynamics of the input process, a generalization of the standard theory of stochastic realization is necessary.

We shall call a model of the type

$$x(t+1) = Ax(t) + Bu(t), \tag{4.1a}$$

$$y_u(t) = Cx(t) + Du(t), \tag{4.1b}$$

a *deterministic realization of y_u in terms of u*. Models of this kind in general have a state space of smaller dimension than those obtained by the standard (Markovian) stochastic realization procedure unless, of course, u is white noise. As usual a realization is called *minimal* if the dimension of the state vector is as small as possible. For minimal realizations it must necessarily hold that (A, B, C) is a minimal triplet. If A has all eigenvalues inside the unit circle ($|\lambda(A)| < 1$), both $x(t)$ and $y_u(t)$ can be expressed as functionals of the infinite past of u. For this reason realizations with this property will be called *causal*.

In order to discuss geometric realization of y_u in terms of u we shall now introduce a technical assumption of 'sufficient richness' of the input process which could probably be relaxed considerably without essentially affecting the validity of the results to be given in this section (except for some lack of uniqueness in the representation formulas). The assumption will permit however a straightforward derivation of the state-space formulas with a minimum amount of technical complications.

Assumption 4.1. For each t the input space $\mathscr{U}$ admits the direct sum decomposition

$$\mathscr{U} = \mathscr{U}_t^- + \mathscr{U}_t^+. \tag{4.2}$$

An analogous condition (namely $\mathscr{U}_t^- \cap \mathscr{U}_t^+ = 0$) is discussed in [16] where it is shown that it is equivalent to strict positivity of the spectral density matrix of u on the unit circle, i.e. $\Phi_u(e^{j\omega}) > cI$, $c > 0$, or to all canonical angles between the past and future subspaces of u being strictly positive (or, in turn, to all canonical correlation coefficients between past and future of the input process being strictly less than one). A slightly stronger version of this condition is found in [26, Chapter II, Section 7].

The *oblique* projection of a random variable $\eta \in \mathscr{U}$ onto $\mathscr{U}_t^-$ along $\mathscr{U}_t^+$ will be denoted by $E_{\|\mathscr{U}_t^+}[\eta|\mathscr{U}_t^-]$. Clearly, if u is a white-noise process, this is the ordinary orthogonal projection onto $\mathscr{U}_t^-$.

The following two lemmas are instrumental for the basic geometric realization procedure presented below.

Lemma 4.1. *Under Assumption* 4.1 *there is a unique transfer function* $\hat{W}(z) = \sum_0^{+\infty} W_k z^{-k}$, *analytic in* $\{|z| > 1\}$, *such that*

$$y_u(t) = \sum_{-\infty}^{t} W_{t-k} u(k). \tag{4.3}$$

Proof. Clearly by Assumption 4.1 u is a purely non-deterministic process and since the past of y_u is included in that of u, y_u is also purely non-deterministic. It follows that $\hat{W}(z)$ is just the transfer function of the Wiener filter $y_u(t) = E[y(t)|\mathscr{U}_{t+1}^-]$, namely,

$$\hat{W}(z) = [\Phi_{yu}(z) G(1/z)^{-T}]_+ G(z)^{-1},$$

where $G(z)$ is the outer (or minimum-phase) spectral factor of Φ_u and the symbol $[\cdot]_+$ means 'analytic part', see e.g. [26, Chapter II]. It is evident that $\hat{W}(z)$ is analytic and, because of non-singularity of Φ_u on the unit circle, unique almost everywhere. □

We shall call a subspace $\mathscr{X}_t \subset \mathscr{U}_t^-$ an *oblique splitting subspace* for the pair $(\hat{\mathscr{Y}}_t^+, \mathscr{U}_t^-)$ if

$$E_{\|\mathscr{U}_t^+}[y_u(t+k) \,|\, \mathscr{U}_t^-] = E_{\|\mathscr{U}_t^+}[y_u(t+k) \,|\, \mathscr{X}_t] \tag{4.4}$$

for all $k \geqslant 0$ and $t \in \mathbb{Z}$. Note that we could substitute y for y_u in these expressions. The *oblique predictor space* $\mathscr{X}_t^{+/-} := E_{\|\mathscr{U}_t^+}[\hat{\mathscr{Y}}_t^+ \,|\, \mathscr{U}_t^-]$ is obviously contained in $\mathscr{U}_t^-$ and is oblique splitting. We shall see shortly that it is in fact a *minimal* oblique splitting subspace.

Define the *extended future space* $\bar{\mathscr{Y}}_t^+ := \hat{\mathscr{Y}}_t^+ \vee \mathscr{X}_t^{+/-}$.

Lemma 4.2. *We have*

$$\mathscr{X}_t^{+/-} = \bar{\mathscr{Y}}_t^+ \cap \mathscr{U}_t^-, \tag{4.5}$$

moreover the direct sum decomposition

$$\bar{\mathscr{Y}}_t^+ = (\bar{\mathscr{Y}}_t^+ \cap \mathscr{U}_t^-) + (\bar{\mathscr{Y}}_t^+ \cap \mathscr{U}_t^+) \tag{4.6}$$

holds for all $t \in \mathbb{Z}$.

This intersection representation extends an analogous result known for 'orthogonal' splitting subspaces

[13, 14]. As shown in Appendix A, it actually holds in greater generality for an arbitrary causal oblique splitting subspace (and even for non-causal ones, provided the input space is also properly extended as in [13, 14]). Since the proof is a bit technical it will be deferred to Appendix A.

The following argument shows how state-space realizations are constructed by a geometric procedure based on the properties of oblique splitting subspaces.

Denote by $\mathscr{U}_t$ the p-dimensional subspace of $\mathscr{U}_t^+$ spanned by the components of $u(t)$. By Assumption 4.1,

$$\mathscr{U}_{t+1}^- = \mathscr{U}_t^- + \mathscr{U}_t$$

and by Lemma 4.2 we can then write

$$\bar{\mathscr{Y}}_{t+1}^+ \cap \mathscr{U}_{t+1}^- = (\bar{\mathscr{Y}}_{t+1}^+ \cap \mathscr{U}_t^-) + (\mathscr{Y}_{t+1}^+ \cap \mathscr{U}_t). \tag{4.7}$$

Now pick a basis vector $x_u(t)$, say of dimension[3] n in $\mathscr{X}_t^{+/-}$ and let $x_u(t+1)$ be the corresponding vector shifted by one unit of time. The n scalar components of $x_u(t+1)$ span $\bar{\mathscr{Y}}_{t+1}^+ \cap \mathscr{U}_{t+1}^-$ and, since $\bar{\mathscr{Y}}_{t+1}^+ \subset \bar{\mathscr{Y}}_t^+$ (this condition is equivalent to markovianness of the splitting subspace in the classical geometric theory, see e.g. [13, Theorem 3.2]) we have

$$(\bar{\mathscr{Y}}_{t+1}^+ \cap \mathscr{U}_t^-) \subset \mathscr{X}_t,$$

so by projecting $x_u(t+1)$ onto the two components of the direct sum decomposition (4.7) we obtain a unique representation of the type

$$x_u(t+1) = A_u x_u(t) + B_u u(t).$$

Similarly, since by the feedback-free property $y_u(t) \in \mathscr{U}_{t+1}^-$, we have

$$y_u(t) \in \bar{\mathscr{Y}}_t^+ \cap \mathscr{U}_{t+1}^- = (\bar{\mathscr{Y}}_t^+ \cap \mathscr{U}_t^-) + (\bar{\mathscr{Y}}_t^+ \cap \mathscr{U}_t),$$

and by projecting $y_u(t)$ onto the two components of the direct sum above we immediately obtain the state-output equation

$$y_u(t) = C_u x_u(t) + D_u u(t).$$

Note that the state process $x_u(t)$ of the representation constructed in this way is stationary by construction and purely non-deterministic as the past $\mathscr{X}_t^-$ is contained in $\mathscr{U}_t^-$. In fact, since $x_u(t)$ is, like $y_u(t)$, a functional of the past history $\mathscr{U}_t$, by Lemma 4.1 applied to the process x_u, it follows that there must be an $m \times p$ matrix function $F(z)$ analytic in $\{|z| > 1\}$, with rows in the space $L_p^2[\Phi_u \, d\omega/2\pi]$ of functions square integrable on the unit circle with respect to the matrix measure $\Phi_u(e^{j\omega}) \, d\omega/2\pi$, such that

$$x_u(t) = \int_{-\pi}^{+\pi} e^{j\omega t} F(e^{j\omega}) \, d\hat{u}.$$

Here $\hat{u}$ denotes the Fourier transform (random orthogonal measure) of the process u. Now, by substituting this into the state equation for x_u derived above, we see, by uniqueness of the spectral representation, that $F(z) = (zI - A_u)^{-1} B_u$. Note that $F(z)$ is rational and actually analytic also on the unit circle since poles of modulus 1 would prevent integrability of the spectrum of x_u, $F(z)\Phi_u(z)F(1/z)'$, on the unit circle. On the other hand, given that $\Phi_u(z)$ has no zeros on the unit circle, there cannot be cancellations with the zeros of $\Phi_u(z)$ either. If the pair (A_u, B_u) is reachable, one easily deduces from the analiticity of $F(z)$ that the eigenvalues of A_u must lie inside the unit disk. Hence, $|\lambda(A_u)| < 1$.

Theorem 4.1. *Assume the joint spectral density of y and u is rational and that the input process satisfies condition* (4.1). *Then the oblique predictor subspace $\mathscr{X}_t^{+/-}$ has finite dimension $\hat{n}$ and*

(1) *For any choice of basis vector $x_u(t)$ in $\mathscr{X}_t^{+/-}$, there are unique matrices (A_u, B_u, C_u, D_u) with (A_u, B_u, C_u) a minimal triplet, such that*

$$x_u(t+1) = A_u x_u(t) + B_u u(t), \tag{4.8a}$$

$$y_u(t) = C_u x_u(t) + D_u u(t). \tag{4.8b}$$

Moreover, the realization (4.8) *is causal, i.e.* $|\lambda(A)| < 1$.

(2) *The state space of any other causal realization of y_u in terms of u has dimension not smaller than $\hat{n}$.*

In other words, $\mathscr{X}_t^{+/-}$ is a minimal state space for realizing y_u in terms of u. In fact, there is a unique (modulo similarity) minimal causal state-space realization of y_u in terms of u and its state space is precisely $\mathscr{X}_t^{+/-}$.

[3] Here for the sake of illustration we assume that $\mathscr{X}_t^{+/-}$ is finite-dimensional.

Proof. That rationality implies existence of finite-dimensional realizations is too well-known to need a proof. In this case the statement follows trivially from the fact that the transfer function $\hat{W}$ is rational and hence has finite-dimensional deterministic realizations. Any such realization with input the process u can be interpreted as a realization of y_u and without loss of generality, the realization can be reduced to be causal, i.e. with $|\lambda(A)| < 1$. The representation (4.8) has been derived above for an arbitrary splitting subspace. Minimality follows from the following argument.

Assume that we have an arbitrary causal state-space representation of y_u with state process x of dimension n and system matrices (A,B,C,D). From this model we obtain

$$y_u(t+k) = \sum_{-\infty}^{t-1} CA^{t+k-s}Bu(s) + \sum_{t}^{t+k} CA^{t+k-s}Bu(s) \quad (4.9a)$$

$$:= y_u(t+k)^- + y_u(t+k)^+ \quad (4.9b)$$

$$= CA^k x(t) + \sum_{t}^{t+k} CA^{t+k-s}Bu(s), \quad k \geqslant 0. \quad (4.9c)$$

Now $y_u(t+k)^-$ and $y_u(t+k)^+$ coincide with the (unique) oblique projections of $y_u(t+k)$ onto the direct summands in the decomposition (4.6). In fact, since the random variables $\{y_u(t+k), k \geqslant 0\}$ generate the future $\hat{\mathscr{Y}}_t^+$, the projections $\{y_u(t+k)^-, k \geqslant 0\}$ generate exactly $\mathscr{X}_t^{+/-}$.

From Eq. (4.9c) in the last line above however it is seen that this subspace is equal to the space spanned by the (scalar components of) $\{CA^k x(t), k \geqslant 0\}$ which are in turn contained in $\mathscr{X}_t := \text{span}\{x_k(t), k \geqslant 0\}$. It follows that $\mathscr{X}_t \supset \mathscr{X}_t^{+/-}$ and hence $\dim x \geqslant \dim x_u$.

Note that the last argument given above shows that the state space $\mathscr{X}_t$ of any causal realization of y_u in terms of u must contain $\mathscr{X}_t^{+/-}$ as a subspace. Therefore, $\mathscr{X}_t^{+/-}$ is just *the unique mimimal state space* for y_u contained in the past of u as claimed. □

5. The joint model

From what has been said in the previous sections, it is clear that a general state-space description for the process y can be obtained by just combining the two separate state-space models for y_s and y_u. For example, the (forward) innovation representation is obtained by combining together (3.7) and (4.8), so that the 'input-output' innovation model of y is

$$\begin{bmatrix} x_s(t+1) \\ x_u(t+1) \end{bmatrix} = \begin{bmatrix} A_s & 0 \\ 0 & A_u \end{bmatrix} \begin{bmatrix} x_s(t) \\ x_u(t) \end{bmatrix} + \begin{bmatrix} 0 \\ B_u \end{bmatrix} u(t) + \begin{bmatrix} B_s \\ 0 \end{bmatrix} e_s(t), \quad (5.1a)$$

$$y(t) = \begin{bmatrix} C_s & C_u \end{bmatrix} \begin{bmatrix} x_s(t) \\ x_u(t) \end{bmatrix} + D_u u(t) + e_s(t). \quad (5.1b)$$

In this model the matrix $\left[\begin{smallmatrix} B_s \\ 0 \end{smallmatrix}\right]$ has the meaning of a steady-state Kalman filter gain and its particular block structure is just a consequence of the fact that the 'deterministic' state subspace of the system (i.e. the subspace where only the last $\hat{n}$ state coordinates are non-zero) is unreachable for the driving white-noise process e_s. This is equivalent to the kernel of the steady-state error-covariance matrix being composed of vectors of the form $\left[\begin{smallmatrix} 0 \\ \hat{x} \end{smallmatrix}\right]$.

More generally, y can be described by a state-space model with the same parallel structure of (5.1) but with an arbitrary minimal stochastic state-space $\tilde{X}_t$ in place of the predictor space $\tilde{X}_t^{+/-}$. As we have seen above there is just *one* minimal state space for the deterministic part of y.

Models of this kind are naturally interpreted as state-space realizations of the familiar ARMAX-type 'input–output' relations $y = W(z)u + G(z)e$ (here we have $W(z) = D_u + C_u(zI - A_u)^{-1}B_u$ and $G(z) = D_s + C_s(zI - A_s)^{-1}B_s$) so often used in the identification literature.[4]

It may happen that, even if the realizations of the two subsystems (stochastic and deterministic) are

[4] It should be recalled that models of this type are meaningful only when the mutual correlation structure of u and e is specified explicitely. This is seldom done in practice. Sometimes u is declared to be 'deterministic' which can be interpreted as u and e being completely uncorrelated. This of course corresponds to the feedback-free case that we are treating here.

minimal, the joint model is not, as there may be loss of observability due to the presence of common modes in the dynamics of the two subsystems. This happens when the transfer functions $W(z)$ and $G(z)$ have common poles and common corresponding eigenspaces. The block structure of (5.1) may then give redundant descriptions of the signals in certain particular cases. These cases are highly non-generic in black-box identification and in practice one need not worry about this evenience. On the contrary, in certain problems one has a pirori some knowledge about the way the input or the noise enter in the system and there may be noise effects which one specifically wants to model as being subject to the same dynamics as the input. In these cases there is actually a need to use models which allow for common dynamics.

6. Estimation given a finite data set

So far we have been working assuming that the complete infinite history of the stationary signals $[y, u]$ was available to build the model. If infinitely long sample data of $[y, u]$ were available it would be a relatively simple matter to translate the geometric stochastic realization procedures of the previous sections into 'subspace' identification algorithms. But this is of course never true in practice where we are instead given only a *finite* input–output data set

$$\begin{aligned}&\{u(0), u(1), \ldots, u(t), \ldots, u(T)\},\\ &\{y(0), y(0), \ldots, y(t), \ldots, y(T)\},\end{aligned} \tag{6.1}$$

and we are faced with the problem of recovering a joint stationary model of the type (5.1) from *finite* input–output data. In this section we shall study the problem of recovering the joint model (5.1) given a finite data set (6.1). For clarity we shall first analyze this equation in an ideal probabilistic setup where the available data are random vectors. We shall take up the question of dealing with real (time-series) data in the next section.

The data (6.1) generate the finite history subspaces of second-order random variables,

$$\mathscr{U}_{[0,T]} := \operatorname{span}\{u(t) \mid 0 \leqslant t \leqslant T\},$$
$$\mathscr{Y}_{[0,T]} := \operatorname{span}\{y(t) \mid 0 \leqslant t \leqslant T\}.$$

Denote the orthogonal complement of $\mathscr{U}_{[0,T]}$ in $\mathscr{U}_{[0,T]} \vee \mathscr{Y}_{[0,T]}$ by $\mathscr{U}_{[0,T]}^{\perp}$, so that

$$\mathscr{U}_{[0,T]} \oplus \mathscr{U}_{[0,T]}^{\perp} = \mathscr{U}_{[0,T]} \vee \mathscr{Y}_{[0,T]}.$$

As we shall see later on, the practical computation of $\mathscr{U}_{[0,T]}^{\perp}$ can be done by an LQ factorization of the data matrix generating $\mathscr{U}_{[0,T]} \vee \mathscr{Y}_{[0,T]}$.

Lemma 6.1. *Let y_u be described by the deterministic realization* (4.8). *Then*

$$E[y(t) \mid \mathscr{U}_{[0,T]}] = E[y_u(t) \mid \mathscr{U}_{[0,T]}] := \hat{y}_u(t), \tag{6.2}$$

where $\hat{y}_u(t)$ is described by the same state-space model but started at a different initial state, namely,

$$\hat{x}_u(t+1) = A_u \hat{x}_u(t) + B_u u(t), \tag{6.3a}$$

$$\hat{y}_u(t) = C_u \hat{x}_u(t) + D_u u(t), \tag{6.3b}$$

$$\hat{x}_u(0) = E[x_u(0) \mid \mathscr{U}_{[0,T]}]. \tag{6.3c}$$

Proof. Follows from (2.4) by noting that $y_s(t) \perp \mathscr{U} \supset \mathscr{U}_{[0,T]}$. □

It follows from this lemma that the deterministic part of (5.1), namely the system matrices (A_u, B_u, C_u, D_u) (in a suitable basis) can be identified by using standard deterministic identification procedure, of the type discussed say in [32, 19], based on the data $\{\hat{y}_u(t),\ u(t) \mid t = 1, 2, \ldots, T\}$. Once the system matrices are computed, the estimate of the initial state $\hat{x}_u(0)$ can also be reconstructed.

The identification of the stochastic subsystem involves the projections of the output data onto the complementary subspace $\mathscr{U}_{[0,T]}^{\perp}$, namely,

$$\hat{y}_s(t) := y(t) - E[y(t) \mid \mathscr{U}_{[0,T]}], \quad 0 \leqslant t \leqslant T. \tag{6.4}$$

Note that these random vectors can in principle be computed from the available data, in fact we have

$$\mathscr{U}_{[0,T]}^{\perp} := \operatorname{span}\{\hat{y}_s(t) \mid 0 \leqslant t \leqslant T\}.$$

The following proposition shows how the projected data (6.4) relate to the 'true' stochastic component y_s.

Proposition 6.1. *Let* $\tilde{y}_u(t) := y_u(t) - \hat{y}_u(t)$ *(the 'smoothing error' of* $y_u(t)$*), then*

$$\hat{y}_s(t) = y_s(t) + \tilde{y}_u(t), \quad 0 \leqslant t \leqslant T. \tag{6.5}$$

Proof. The proof is based essentially on the same argument as in Lemma 6.1, for, from (2.4) by noting that $y_s(t) \perp \mathscr{U} \supset \mathscr{U}_{[0,T]}$, one gets $\hat{y}_s(t) := y_s(t) + y_u(t) - E[y_u(t) \mid \mathscr{U}_{[0,T]}] = y_s(t) + \tilde{y}_u(t)$. □

Hence for finite data length, the projection $\hat{y}_s(t)$ of the output on the complementary subspace $\mathscr{U}_{[0,T]}^{\perp}$ does not coincide with the stochastic component $y_s(t)$, as it would instead have happened for *infinite* data length. The 'ideal' projection $y_s(t)$ is affected by an additional 'smoothing error' term $\tilde{y}_u(t)$ which depends on the error on the estimate of the initial state of the deterministic component, $\tilde{x}_u(0) := x_u(0) - \hat{x}_u(0)$. It has the form

$$\tilde{y}_u(t) = C_u A_u^t \tilde{x}_u(0), \quad 0 \leqslant t \leqslant T.$$

This additional term is a source of difficulty in identification of the stochastic part since if it is neglected it tends to produce a stochastic model of y_s of a much higher dimension than the true order $\tilde{n}$. Therefore, a preliminary step necessary for the identification of the stochastic component is to filter it out somehow.

A detailed description of this prefiltering step would unfortunately require too much of a diversion from the main theme of this paper and we shall not present it here. For a succinct discussion we shall refer the reader to the forthcoming paper [23].

7. How to deal with real data

In practice, instead of random variables one has just a collection of input–output data,

$$\{u_0, u_1, \ldots, u_t, \ldots, u_N\}, \qquad \{y_0, y_1, \ldots, y_t, \ldots, y_N\}, \tag{7.1}$$

with $u_t \in \mathbb{R}^p$, $y_t \in \mathbb{R}^m$, measured during an experiment. We shall assume that the sample size N is very large and that the data have been preprocessed so as to be compatible with the basic assumption of (wide-sense) stationarity and zero mean of the previous section. More specifically, we shall assume that we can pick M large enough so that the time averages

$$\frac{1}{M+1} \sum_{t=t_0}^{M+t_0} \begin{bmatrix} u_{t+\tau} \\ y_{t+\tau} \end{bmatrix} \begin{bmatrix} u_t \\ y_t \end{bmatrix}', \quad \tau \geqslant 0, \tag{7.2}$$

are practically independent of the initial time t_0 and arbitrarily close to a bona-fide stationary covariance matrix sequence

$$\Lambda(\tau) := \begin{bmatrix} \Lambda_{uu}(\tau) & \Lambda_{uy}(\tau) \\ \Lambda_{yu}(\tau) & \Lambda_{yy}(\tau) \end{bmatrix}, \quad \tau \geqslant 0. \tag{7.3}$$

Under these assumptions on the data we can translate the stochastic realization constructions of the previous sections into geometric procedures based on manipulations of the observed input–output time series. Although these manipulations are quite standard in the subspace identification literature, most authors seem to be unaware of the fact that there is a precise one-to-one correspondence which automatically translates the 'abstract' stochastic geometrical setting of stochastic realization into the 'concrete' linear-algebraic data processing setup of subspace identification. In this section we shall briefly review the basic ideas behind this correspondence. For clarity of exposition we intially assume that $N = \infty$ (so that the time series are actually doubly infinite).

For each $t \in \mathbb{Z}$ define the $p \times \infty$ and $m \times \infty$ matrices [5]

$$u(t) := [u_t, u_{t+1}, u_{t+2}, \ldots], \tag{7.4a}$$

$$y(t) := [y_t, y_{t+1}, y_{t+2}, \ldots], \tag{7.4b}$$

and consider the sequences $u := \{u(t) \mid t \in \mathbb{Z}\}$ and $y := \{y(t) \mid t \in \mathbb{Z}\}$. These sequences will play a very similar role to the stationary processes u and y of the previous sections.

Define the vector spaces $\mathscr{U}$ and $\mathscr{Y}$ of all finite linear combinations

$$\mathscr{U} := \left\{ \sum a_k' u(t_k) \quad a_k \in \mathbb{R}^p, t_k \in \mathbb{Z} \right\}, \tag{7.5}$$

$$\mathscr{Y} := \left\{ \sum a_k' y(t_k) \quad a_k \in \mathbb{R}^m, t_k \in \mathbb{Z} \right\}. \tag{7.6}$$

Note that the vector spaces $\mathscr{U}$ and $\mathscr{Y}$ are just the row spaces of the two families of semi-infinite matrices (7.4) or, equivalently, of the infinite Hankel

[5] The use of the same symbols used for the random processess u and y in Sections 1 and 2 is intentional.

matrices

$$U_\infty := \begin{bmatrix} \vdots \\ u(t) \\ u(t+1) \\ u(t+2) \\ \vdots \end{bmatrix}, \qquad Y_\infty := \begin{bmatrix} \vdots \\ y(t) \\ y(t+1) \\ y(t+2) \\ \vdots \end{bmatrix},$$

while of course $\mathscr{U} \vee \mathscr{Y}$ can be seen as the row space of the compound matrix $\begin{bmatrix} U_\infty \\ Y_\infty \end{bmatrix}$ constructed by stacking all shifted future tails of the observed data. This vector space of scalar semi-infinite sequences (rows) can be equipped with an inner product, which is first defined on the generators by the bilinear form

$$\left\langle a' \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}, b' \begin{bmatrix} u(j) \\ y(j) \end{bmatrix} \right\rangle := \lim_{N\to\infty} \frac{1}{N+1} \times \sum_{t=0}^{N} a' \begin{bmatrix} u_{t+k} \\ y_{t+k} \end{bmatrix} \begin{bmatrix} u_{t+j} \\ y_{t+j} \end{bmatrix}' b = a' \Lambda(k-j) b, \tag{7.7}$$

and then extended by linearity to all finite linear combinations of elements of $\mathscr{U} \vee \mathscr{Y}$. This inner product is non-degenerate[6] if the block-Toeplitz matrix T_k, constructed from the covariance data $\{\Lambda(0), \Lambda(1), \ldots, \Lambda(k), \ldots\}$, is positive definite for all k [16]. Note also that the limit does not change if in the limits of the sum (7.7) $t = 0$ is replaced by an arbitrary initial instant t_0, so that

$$\left\langle a' \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}, b' \begin{bmatrix} u(j) \\ y(j) \end{bmatrix} \right\rangle = \left\langle a' \begin{bmatrix} u(t_0+k) \\ y(t_0+k) \end{bmatrix}, b' \begin{bmatrix} u(t_0+j) \\ y(t_0+j) \end{bmatrix} \right\rangle$$

for all t_0 (wide-sense stationarity). We also define a *shift operator* σ on the family of semi-infinite matrices (7.4), by setting

$$\sigma a' u(t) = a' u(t+1), \quad t \in \mathbb{Z}, \ a \in \mathbb{R}^p,$$
$$\sigma a' y(t) = a' y(t+1), \quad t \in \mathbb{Z}, \ a \in \mathbb{R}^m,$$

defining a linear map which is isometric with respect to the inner product (7.7) and extendable by linearity to all of $\mathscr{U} \vee \mathscr{Y}$.

[6] This means $\langle \xi, \xi \rangle = 0 \Leftrightarrow \xi = 0$.

By closing the vector space $\mathscr{U} \vee \mathscr{Y}$ with respect to convergence in the norm induced by the inner product (7.7), we obtain a Hilbert space[7] $\overline{\mathscr{U} \vee \mathscr{Y}}$ to which the shift operator σ is extended by continuity as a unitary operator.

As explained in more detail in [16], this Hilbert space framework is isometrically isomorphic to the abstract 'stochastic' geometric setup used in the previous sections. We can formally think of the observed time series (7.1) as ergodic sample paths of two jointly wide-sense stationary stochastic process $\boldsymbol{u}, \boldsymbol{y}$, having joint covariance matrices equal to the limit (7.3) of the sum (7.2) as $M \to \infty$. Then, at least as far as first- and second-order moments are concerned, the tail sequences u and y defined in (7.4) behave exactly like the abstract counterparts $\boldsymbol{u}$ and $\boldsymbol{y}$. In particular, all second-order moments of the two random processes can equivalently be calculated in terms of the tail sequences u and y provided we substitute expectations with ergodic limits of the type (7.7). Since we only worry about second-order properties in this paper, we may even regard the tail sequences u and y of (7.4) as being the same object as the two underlying stochastic process $\boldsymbol{u}$ and $\boldsymbol{y}$. This requires just thinking of 'random variables' as being semi-infinite strings of numbers and defining the expectation of products $E\{\xi\eta\}$ as being the inner product of the corresponding rows ξ and η. For reasons of uniformity of notation the inner product (7.7) will then be denoted

$$\langle \xi, \eta \rangle = E\{\xi\eta\}, \tag{7.8}$$

Here we allow $E\{\cdot\}$ to operate on matrices, taking inner products row by row.

Hence, all results in the geometric theory of stochastic realization can be carried over to the present framework. The orthogonal projection of a row (random variable) ξ onto a subspace $\mathscr{H}$ of the space $\mathscr{U} \vee \mathscr{Y}$ will still be denoted $E[\xi \mid \mathscr{H}]$. Whenever $\mathscr{H}$ is given as the rowspace of some matrix of generators H, we shall write $E[\xi \mid H]$ to denote the projection expressed

[7] Note that the symbols $\mathscr{U}$ and $\mathscr{Y}$ have a different meaning in this section as they just denote real inner-product spaces which need not be closed with respect to the inner product structure defined by (7.7). Since however we will not have any use for the completed spaces in the following, we shall not introduce special symbols for them.

(perhaps non-uniquely) in terms of the generators. It is clear that for finitely generated subspaces we have the representation formula

$$E[\xi|H] = E(\xi H')[E(HH')]^{\#}H, \tag{7.9}$$

and in case of linearly independent rows we can substitute the pseudoinverse # with a true inverse.

8. Identification based on finite data

For data of finite length N the inner product (7.7) is approximated by a finite sum

$$\langle \xi, \eta \rangle \cong \frac{1}{N+1} \sum_{t=0}^{N} \xi_t \eta_t,$$

which makes it essentially the same as the ordinary Euclidean inner product in $\mathbb{R}^N$.

Assume now the integer N is chosen so large that the time averages in the ergodic limit (7.2) are sufficiently close to the true covariance for say $\tau = 0, 1, \ldots, 2k$. Fix a 'present' time $t = k$ and define the four block Hankel matrices of dimension $pk \times (N+1) \ldots mk \times (N+1)$ formed by input and output data as

$$U_k^- = \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(k-1) \end{bmatrix} = \begin{bmatrix} u_0 & u_1 & \cdots & u_N \\ u_1 & u_2 & \cdots & u_{N+1} \\ \vdots & \vdots & & \vdots \\ u_{k-1} & u_k & \cdots & u_{k+N-1} \end{bmatrix}, \tag{8.1}$$

$$Y_k^- = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(k-1) \end{bmatrix} = \begin{bmatrix} y_0 & y_1 & \cdots & y_N \\ y_1 & y_2 & \cdots & y_{N+1} \\ \vdots & \vdots & & \vdots \\ y_{k-1} & y_k & \cdots & y_{k+N-1} \end{bmatrix}, \tag{8.2}$$

$$U_k^+ = \begin{bmatrix} u(k) \\ u(k+1) \\ \vdots \\ u(2k-1) \end{bmatrix} = \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+N} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+N+1} \\ \vdots & \vdots & & \vdots \\ u_{2k-1} & u_{2k} & \cdots & u_{2k+N-1} \end{bmatrix}, \tag{8.3}$$

$$Y_k^+ = \begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(2k-1) \end{bmatrix} = \begin{bmatrix} y_k & y_{k+1} & \cdots & y_{k+N} \\ y_{k+1} & y_{k+2} & \cdots & y_{k+N+1} \\ \vdots & \vdots & & \vdots \\ y_{2k-1} & y_{2k} & \cdots & y_{2k+N-1} \end{bmatrix}. \tag{8.4}$$

The relative rowspaces $\mathscr{U}_k^-$ ($\mathscr{Y}_k^-$) are the past input (output) spaces generated by the rows of the $p \times (N+1)$ tail matrices $u(t)$ (respectively $m \times (N+1)$ matrices $y(t)$) for $0 \leqslant t < k$, while $\mathscr{U}_k^+$ and $\mathscr{Y}_k^+$ are the future inputs (outputs) spaces spanned by $u(t)$ (and $y(t)$) for $k \leqslant t < 2k$. Since the tail matrix sequences we can form with the observed data are necessarily finite, these vector spaces can describe in reality only *finite* past and future histories of the data at time k. For simplicity of notations we use symbols that are not informative of this fact.[8]

Let us define also the stacked $2k \times (N+1)$ block Hankel matrices

$$U := \begin{bmatrix} U_k^- \\ U_k^+ \end{bmatrix}, \qquad Y := \begin{bmatrix} Y_k^- \\ Y_k^+ \end{bmatrix},$$

the relative rowspaces $\mathscr{U}_{[0,2k-1]}$ and $\mathscr{Y}_{[0,2k-1]}$ will also be denoted by the shorthands $\mathscr{U}$ and $\mathscr{Y}$. This should cause no confusion with the infinite data subspaces introduced in Section 2 since from now on only *finite* data set are used.

The first step of the identification algorithm is to separate the (finite-history) 'deterministic' and 'stochastic' components of the output. This is done

[8] More accurate notations would be $\mathscr{U}_k^- := \mathscr{U}_{[0,k)}$, $\mathscr{Y}_k^- := \mathscr{Y}_{[0,k)}$, $\mathscr{U}_k^+ := \mathscr{U}_{[k,2k)}$, $\mathscr{Y}_k^+ := \mathscr{Y}_{[k,2k)}$.

by computing

$$\hat{y}_u(t) = E[y(t)|\mathscr{U}],$$
$$\hat{y}_s(t) = E[y(t)|\mathscr{U}^\perp], \quad t = 0,\dots,2k-1, \tag{8.5}$$

where the orthogonal projections are with respect to the Euclidean inner product defined above; compare with (6.2) and (6.4).

For the actual computation of the deterministic and the stochastic components of y we use the LQ factorization

$$\begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} L_{uu} & 0 \\ L_{yu} & L_{yy} \end{bmatrix} \begin{bmatrix} Q_u^T \\ Q_y^T \end{bmatrix}, \tag{8.6}$$

where $Q_u^T Q_u = I$, $Q_y^T Q_y = I$, $Q_u^T Q_y = 0$ and L_{uu}, L_{yy} are lower triangular. The LQ factorization is a key step in many subspace identification algorithms.

Lemma 8.1. *The Hankel matrices of the deterministic and stochastic components of y are obtained from the LQ decomposition above, respectively, as*

$$\hat{Y}_u = E[Y|\mathscr{U}] = YQ_uQ_u^T = L_{yu}Q_u^T$$

and

$$\hat{Y}_s = E[Y|\mathscr{U}^\perp] = YQ_yQ_y^T = L_{yy}Q_y^T.$$

Proof. This follows immediately by noting that the rows of Q_u^T form an orthonormal basis for the rowspace of U and those of Q_y^T an orthonormal basis for the orthogonal complement $\mathscr{U}^\perp$ in $\mathscr{U} \vee \mathscr{Y}$. □

8.1. Identification of the deterministic part

Standard subspace methods, e.g. [19], to identify a stationary state-space model for the deterministic part y_u, are based on computing the state $\hat{x}_u(t)$ of the system (6.3) at two time instants, say k and $k+1$ and then solve in the least-squares sense,

$$\begin{bmatrix} \hat{x}_u(k+1) \\ \hat{y}_u(k) \end{bmatrix} = \begin{bmatrix} A_u & B_u \\ C_u & D_u \end{bmatrix} \begin{bmatrix} \hat{x}_u(k) \\ u(k) \end{bmatrix}, \tag{8.7}$$

where $\hat{x}_u(k) := [(x_u)_k, \dots, (x_u)_{k+N-1}]$ is a basis for the state space at time k.

Theoretically (see Theorem 4.1) the state $x_u(k)$ is a basis of the intersection of the future rowspan $\bar{\mathscr{Y}}_u^+$ with the *infinite past space* of u at time k. Instead we only have the finite history U^- available (here to simplify notations we drop the subscript k).

For a finite data set, the state $x_u(k)$ is, strictly speaking, no longer a basis for the intersection $\bar{\mathscr{Y}}_u^+ \cap \mathscr{U}^-$ because of the presence of a generally non-zero initial condition $x_u(0)$ at the initial data point $t = 0$. It can however be shown by using arguments similar to [19, Theorem 3, pp. 223–224], that for k large enough, we have

$$\hat{x}_u(k) = \text{ basis in } (\hat{\mathscr{Y}}_u^+ \vee \mathscr{U}^+) \cap \mathscr{U}^-, \tag{8.8}$$

so that we are still led to compute a basis for the intersection of two larger subspaces of the same structure.

Below we give an algorithm to compute a well-conditioned basis for the intersection of two subspaces of the form $\bar{\mathscr{Y}}_u^+ \cap \mathscr{U}^-$, based on the GSVD [21]. Note that to determine the unknown system matrices in (8.7), we need to apply the basis selection algorithm at two consecutive time instants. In fact, the basis at time $k+1$ must be chosen in such a way that it corresponds to the basis at time k, $x_u(k)$, stationarily time-shifted by one time unit, otherwise we would not get the right constant parameters in the realization. This means that after applying the algorithm at time $k+1$, a suitable change of basis is necessary.

8.2. Algorithm for computing a basis for the subspace $\hat{\mathscr{X}} := \bar{\mathscr{Y}}^+ \cap \mathscr{U}^-$ based on the GSVD

We employ the canonical correlation approach described in [9] and numerically determine the dimension $\hat{n}$ of $\hat{\mathscr{X}}$.

We first introduce the LQ factorization

$$\begin{bmatrix} U^- \\ \bar{Y}^+ \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix},$$

where we assume that $[L_{21}\ L_{22}]$ has full row rank. Then there exist $V(kp \times kp)$, $Z(kp \times kp)$ and a non-singular $X(km \times km)$ such that [21]

$$L_{21} = X\hat{C}V^T = X\,\mathrm{diag}(c_1,\dots,c_v)V^T,$$
$$L_{22} = X\hat{S}Z^T = X\,\mathrm{diag}(\hat{s}_1,\dots,\hat{s}_v)Z^T,$$

where $c_1 \geqslant \cdots \geqslant c_v \geqslant 0$, $\hat{s}_v \geqslant \cdots \geqslant \hat{s}_1 \geqslant 0$, $\hat{C}\hat{C}^T + \hat{S}\hat{S}^T = I_{km}$ and $v := \min(pk, mk)$.

Letting

$$\hat{Q}^T := \hat{C}V^TQ_1^T + \hat{S}Z^TQ_2^T,$$

we have $\bar{Y}^+ = X\hat{Q}^T$ and $\hat{Q}^T\hat{Q} = I_{kp}$, so that the rows of Q_1^T and $\hat{Q}^T$ form orthogonal bases for the rowspaces U^- and $\bar{Y}^+$, respectively. Moreover, we have $c_i = \cos(\theta_i)$, $i = 1,\dots,\nu$, where θ_i is the ith smallest principal angle between the rowspaces of $\bar{Y}^+$ and U^-.

If $1 = \cos(\theta_1) = \cdots = \cos(\theta_{\hat{n}}) > \cos(\theta_{\hat{n}+1})$ for some $\hat{n}$, the first $\hat{n}$ rows of V^TQ_1 span the intersection $\bar{Y}_t^+ \cap U_t^-$ [9]. Since this does rarely happen in actual computations, in practice we determine $\hat{n}$ by selecting the first $\hat{n}$ principal angles which are 'numerically' equal to zero.

Note that the Hankel matrix $\hat{H}$ of the deterministic system is

$$\hat{H} = \frac{1}{N}\hat{Y}^+(U^-)^T = \frac{1}{N}X\hat{C}V^TL_{11}^T.$$

Let $\hat{L}_- := L_{11}/\sqrt{N}$, $\hat{L}_+ := X/\sqrt{N}$, where $\hat{L}_-, \hat{L}_+$ are invertible. Hence, we get

$$\hat{L}_+\hat{H}_{21}L_-^{-T} = \hat{C}V^T = \hat{U}_1\hat{\Sigma}V_1^T, \tag{8.9}$$

where $\hat{\Sigma} = \operatorname{diag}(c_1,\dots,c_{\hat{n}})$, $\hat{U}_1 = \begin{bmatrix} I_{\hat{n}} \\ 0 \end{bmatrix}$, and $\hat{n}$ denotes the number of $\hat{c}_i$'s that are effectively unity. Thus, we see that (8.9) is a SVD factorization of the normalized Hankel matrix, so that the corresponding state vector at time k is

$$x_u(k) = \hat{\Sigma}^{1/2}V_1\hat{L}_-^{-1}U_k^-.$$

To compute the state vector $x_u(k+1)$, we use a similar procedure and then rescale by a non-singular matrix so as to force the observability matrices to coincide. This is also done in the 'stochastic' subspace algorithm of [29].

8.3. *Identification of the stochastic part*

The identification of the stochastic subsystem (3.7) is done by processing the 'stochastic' data matrix $\hat{Y}_s$ computed in the previous step. A prefiltering algorithm on the data $\hat{Y}_s$ to eliminate the deterministic 'smoothing error' component is discussed in [23]. Once the 'theoretical' stochastic component is obtained, the identification of the stochastic subsytem can be done by the techniques described, e.g., in [3, 5, 29, 16]. As explained in [16], identification of the model in a *stochastically balanced* canonical form is to be recommended in order to guarantee positivity after the unavoidable truncation step which retains only the significantly larger singular values. The state (basis) vector in the predictor space should then be chosen to be the corresponding truncated subvector of canonical variates.

9. Conclusions

In this paper, we have solved the stochastic realization problem with exogenous inputs in the absence of feedback. We have introduced a class of dynamical models of a very simple structure where the stochastic and deterministic dynamics of the output process are completely decoupled. These models are generically minimal and can be advantageously used for identification. In this setting the identification with exogenous inputs can be split into two separate subproblems of identifying the deterministic component and then of identifying the stochastic component. The second identification step requires prefiltering of the data by means of a filter constructed from the deterministic model.

Acknowledgements

We would like to thank the referee for the careful review and for some useful suggestions which have led to a considerable improvement of this paper.

Appendix A. Proof of Lemma 4.2

We shall fix $t = 0$ and drop the subscripts throughout. The proof will be done in several steps.

In order to streamline notations, assume here that y and u are jointly stationary p.n.d. processes of dimensions m and p, u has a strictly positive spectral density and let $\mathscr{Y}^- \subset \mathscr{U}^-$, i.e. y admits a causal convolution representation in terms of u of the form (4.3). Write

$$y(t) = (H_W u)(t) + (W^+u)(t), \tag{A.1}$$

where

$$(H_W u)(t) := \sum_{-\infty}^{-1} W_{t-k} u(k),$$
$$(W^+ u)(t) := \sum_{0}^{t} W_{t-k} u(k). \tag{A.2}$$

Obviously, $(H_W u)(t) \in \mathscr{U}^-$ and $(W^+ u)(t) \in \mathscr{U}^+$ for $t \geq 0$. This fact is formally recorded in the following lemma.

Lemma A.1. *For $t \geq 0$ the random variable $(H_W u)(t)$ is the oblique projection of $y(t)$ onto $\mathscr{U}^-$ and belongs to $\mathscr{X}^{+/-} := E_{\|\mathscr{U}^+}[\mathscr{Y}^+|\mathscr{U}^-]$. Hence, it belongs also to all causal oblique splitting subspaces $\mathscr{X}$.*

The last statement of this lemma is an immediate consequence of the minimality of $\mathscr{X}^{+/-}$ among all causal splitting subspaces.

Lemma A.2. *Let $\mathscr{X}$ be an oblique splitting subspace and define*

$$\bar{\mathscr{S}} := \mathscr{Y}^+ \vee \mathscr{X}.$$

Then $\mathscr{X}$ is the minimal oblique splitting subspace for $\bar{\mathscr{S}}$ and $\mathscr{U}^-$ contained in $\mathscr{U}^-$.

Proof. Since every element $\bar{s}$ of $\bar{\mathscr{S}}$ has the form $\bar{s} = y^+ + x$, $y^+ \in \mathscr{Y}^+$, $x \in \mathscr{X}$ and $E_{\|\mathscr{U}^+}[y^+|\mathscr{U}^-] \in \mathscr{X}^{+/-} \subset \mathscr{X} \subset \mathscr{U}^-$,

$$\overline{\text{span}}\{E_{\|\mathscr{U}^+}[\bar{s}|\mathscr{U}^-] \,|\, \bar{s} \in \bar{\mathscr{S}}\} = \mathscr{X}. \qquad \square$$

Formula (4.5) is a particular case of the following result.

Lemma A.3. *Let the symbols have the same meaning as in Lemma* A.2. *Then*

$$\bar{\mathscr{S}} \cap \mathscr{U}^- = \mathscr{X}.$$

Proof. First note that $\mathscr{X}$ contains the intersection $\bar{\mathscr{S}} \cap \mathscr{U}^-$. For if $\eta \in \bar{\mathscr{S}} \cap \mathscr{U}^-$ then clearly it belongs to $E_{\|\mathscr{U}^+}[\bar{\mathscr{S}}|\mathscr{U}^-]$ which is equal to $\mathscr{X}$ in force of the previous lemma.

Then just observe that, conversely, the intersection contains $\mathscr{X}$, since $\bar{\mathscr{S}} \supset \mathscr{X}$ and $\mathscr{U}^- \supset \mathscr{X}$. This proves the lemma. $\square$

The result in particular applies to the extended future space $\bar{\mathscr{Y}}^+ = \mathscr{Y}^+ \vee \mathscr{X}^{+/-}$ (this is in a sense the 'minimal' $\bar{\mathscr{S}}$). In general, it is *not true* that $\mathscr{X}^{+/-} = \mathscr{Y}^+ \cap \mathscr{U}^-$ as

$$E_{\|\mathscr{U}^+}[\mathscr{Y}^+|\mathscr{U}^-] \supset \mathscr{Y}^+ \cap \mathscr{U}^-$$

properly, unless some special conditions are satisfied.

We now prove the identity (4.6) in the following lemma.

Lemma A.4. *Let the symbols have the same meaning as in Lemma* A.2. *Then*

$$\bar{\mathscr{S}} = (\bar{\mathscr{S}} \cap \mathscr{U}^-) + (\bar{\mathscr{S}} \cap \mathscr{U}^+). \tag{A.3}$$

Proof. That

$$\bar{\mathscr{S}} \supset (\bar{\mathscr{S}} \cap \mathscr{U}^-) + (\bar{\mathscr{S}} \cap \mathscr{U}^+) \tag{A.4}$$

is obvious since both terms on the right-hand side are subspaces of $\bar{\mathscr{S}}$. We shall show that the opposite inclusion also holds. In effect, decomposing y_u as in (A.1), i.e. $y_u(t) = (H_W u)(t) + (W^+ u)(t)$, from Lemma A.1, we have $(H_W u)(t) \in \mathscr{X} \subset \bar{\mathscr{S}}$ for $t \geq 0$, so that for $t \geq 0$ necessarily $(W^+ u)(t) = y_u(t) - (H_W u)(t) \in \bar{\mathscr{S}}$ as well. In fact, $(H_W u)(t) \in \bar{\mathscr{S}} \cap \mathscr{U}^-$ and $(W^+ u)(t) \in \bar{\mathscr{S}} \cap \mathscr{U}^+$. Taking finite linear combinations of the form $\sum a_k' y_u(t_k)$, $a_k \in \mathbb{R}^m$, $t_k \geq 0$ and then closing in the Hilbert space norm of second-order random variables gives immediately the opposite inclusion to (A.4). $\square$

References

[1] H. Akaike, "Canonical correlation analysis of time series and the use of an information criterion, in: R. Mehra and D. Lainiotis, eds., *System Identification: Advances and Case Studies*, Academic Press, New York, 1976, pp. 27–96.

[2] B.D.O. Anderson and M.R. Gevers, "Identifiability of linear stochastic systems operating under linear feedback", *Automatica*, Vol. 18, 1982, pp. 195–213.

[3] M. Aoki, *State Space Modeling of Time Series*, 2nd Edition, Springer, Berlin, 1990.

[4] P.E. Caines and C.W. Chan, "Estimation, identification and feedback", in: R. Mehra and D. Lainiotis, eds., *System Identification: Advances and Case Studies*, Academic Press, New York, 1976.

[5] U.B. Desai, D. Pal and R.D. Kirkpatrick, "A realization approach to stochastic model reduction", *Internat. J. Control*, Vol. 42, 1985, pp. 821–838.

[6] P.A. Fuhrmann, *Linear Operars and Systems in Hilbert Space*, McGraw-Hill, New York, 1981.

[7] M.R. Gevers and B.D.O. Anderson, "Representation of jointly stationary feedback free processes", *Internat. J. Control*, Vol. 33, 1981, pp. 777–809.

[8] M.R. Gevers and B.D.O. Anderson, "On jointly stationary feedback free stochastic processes", *IEEE Trans. Automat. Control*, Vol. AC-27, 1982, pp. 431–436.

[9] G.H. Golub and C.R. Van Loan, *Matrix Computations*, 2nd Edition, The Johns Hopkins Univ. Press, Baltimore, MA, 1989.

[10] C.W.J. Granger, "Economic processes involving feedback", *Inform. Control*, Vol. 6, 1963, pp. 28–48.

[11] H. Helson, *Lectures on Invariant Subspaces*, Academic Press, New York, 1976.

[12] W.E. Larimore, "Canonical variate analysis in identification, filtering, and adaptive control", *Proc. 29th CDC*, 1990, pp. 596–604.

[13] A. Lindquist and G. Picci, "Realization theory for multivariate stationary Gaussian processes", *SIAM J. Control Optimiz.*, Vol. 23, 1985, pp. 809–857.

[14] A. Lindquist and G. Picci, "A geometric approach to modelling and estimation of linear stochastic systems", *J. Math. Systems, Estimation Control*, Vol. 1, 1991, pp. 241–333.

[15] A. Lindquist and G. Picci, "On subspace-methods identification and stochastic model reduction", *Proc. 10th IFAC Symp. on System Identification*, Copenhagen, DK, Vol. 2, 1994, pp. 397–403.

[16] A. Lindquist and G. Picci, "Canonical correlation analysis approximate covariance extension and identification of stationary time series", *Automatica*, Vol. 32, No. 5, 1996, pp. 709–733.

[17] A. Lindquist, G. Picci and G. Ruckebusch, "On minimal splitting subspaces and Markovian representation", *Math. System Theory*, Vol. 12, 1979, pp. 271–279.

[18] L. Ljung, *System Identification–Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[19] M. Moonen, B. De Moor, L. Vanderberghe and J. Vandewalle, "On- and off-line identification of linear state-space models", *Internat. J. Control*, Vol. 49, 1989, pp. 219–232.

[20] M. Moonen and J. Vandewalle, "QSVD approach to on- and off-line state-space identification", *Internat. J. Control*, Vol. 51, 1990, pp. 1133–1146.

[21] C.C. Paige and M.A. Saunders, "Towards a generalized singular value decomposition", *SIAM J. Numer. Anal.*, Vol. 18, 1981, pp. 398–405.

[22] G. Picci, "Stochastic realization of Gaussian processes", *Proc. IEEE*, Vol. 64, 1976, pp. 112–122.

[23] G. Picci and T. Katayama, "A simple subspace identification algorithm with exogenous inputs", *Proc. 1996 Triennial IFAC Congress*, San Francisco, CA, Paper no. 0916, Session 3a-06-5.

[24] G. Picci and S. Pinzoni, "Acausal models and balanced realizations of stationary processes", *Linear Algebra Appl.*, Vols. 205–206, 1994, pp. 997–1043.

[25] G. Picci and S. Pinzoni, "Dynamic factor-analysis models for stationary processes", *IMA J. Math. Control Inform.*, Vol. 3, No. 2, 1986, pp. 185–210.

[26] N.I. Rozanov, *Stationary Random Processes*, Holden-Day, San Francisco, CA, 1963.

[27] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[28] J.H. van Schuppen, "Stochastic realization of a Gaussian stochastic control system", *Acta Applicandae Methematicae*, Vol. 35, 1994, pp. 193–212.

[29] P. Van Overschee and B. De Moor, "Subspace algorithms for the stochastic identification problem", *Automatica*, Vol. 29, 1993, pp. 649–660.

[30] P. Van Overschee and B. De Moor, N4SID: Subspace algorithms for the identification of combined deterministic–stochastic systems", *Automatica*, Vol. 30, 1994, pp. 75–93.

[31] P. Van Overschee and B. De Moor, "A unifying theorem for subspace system identification algorithms and its interpretation", *Proc. 10th IFAC Symp. on System Identification*, Vol. 2, 1994, pp. 145–156.

[32] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in Innovations form from input–output data", *Automatica*, Vol. 30, 1994, pp. 61–74.

[33] M. Viberg, B. Ottersten, B. Wahlberg and L. Ljung, "A statistical-perspective on state-space modeling using subspace methods", *Proc. 30th CDC*, London, 1991, pp. 1337–1342.

[34] H.P. Zeiger and A.J. McEwen, "Approximate linear realization of given dimension via Ho's algorithm", *IEEE Trans. Automat. Control*, Vol. AC-19, 1974, p. 153.