CWI Quarterly Volume 9 (3) 1996, pp. 205 – 240

Geometric Methods in Stochastic Realization and System Identification

Giorgio Picci
Dipartimento di Elettronica e Informatica
University of Padova and LADSEB-CNR
35131 Padua, Italy
e-mail: picci@dei.unipd.it

In this paper we discuss some recent advances in modeling and identification of stationary processes. We point out that identification of linear state-space models for stationary signals can be seen as stochastic realization of wide-sense stationary processes in an appropriate background Hilbert space. The geometric theory of stochastic realization developed in the last two decades plays an important role in this interpretation. Identification of models with exogenous inputs in conditions of absence of feedback can also be formulated as a stochastic realization problem. We discuss procedures for constructing minimal state-space models in presence of inputs, based on a generalization of stochastic realization theory for time series and we discuss geometric procedures for identifying (generically) minimal state-space models with inputs. This approach leads to numerical linear algebraic algorithms which have been named "subspace methods" in the literature. It has important advantages over the traditional parametric optimization approach, since it attacks directly the dynamic model building problem by system theoretic methods and leads to procedures which are more transparent and more structured than those traditionally used and found in the literature.

1. Introduction

Stochastic Realization theory deals with modeling of random processes. Given a vector (say m-dimensional) process $y = \{y(t)\}$, one wants to find representations of y in terms of simpler and more basic random processes, such as white noise, Markov processes etc. In particular it deals with procedures for constructing models of stationary processes, of the following form

$$\begin{cases} x(t+1) &= Ax(t) + Bw(t) \\ y(t) &= Cx(t) + Dw(t) \end{cases}, \tag{1}$$

where $\{w(t)\}$ is a vector normalized white noise process, i.e. $E\{w(t)w(s)'\} = I\delta(t-s)$, $E\{w(t)\} = 0$, δ being the Kronecker delta function. This representation is called a *(linear) state-space realization* of the process y. It involves auxiliary variables, i.e. random quantities which are not given as a part of the original data, such as the *state process* x (a stationary Markov process) and the *generating white noise* w, whose peculiar properties lead to representations of y by models having the desired structure. Constructing these auxiliary processes is part of the realization problem.

In this paper we shall present a survey of state-space realization of wide-sense stationary second-order processes and discuss applications of the theory to identification. This may look like a very particular modeling problem to deal with, and it should be said that there are other interesting areas of stochastic modeling and other possible viewpoints to discuss the subject, see e.g. [62, 50, 63]. The present choice has been motivated by the relative degree of maturity reached by this research area and by the desire of showing the practical applicability of the ideas involved to an important statistical problem.

One may add that wide-sense linear models like 1 are extremely important in applications for a variety of reasons, including the relative simplicity of the probabilistic treatment and the fact that most of the time in practice only second order statistics are available to describe random phenomena. They are the starting point for popular estimation and control algorithms like Kalman filtering, LQG control etc..

The model 1 is a state-space realization of a single stationary process y [1]. More generally one may want to construct state-space models involving also exogenous (input) variables. The mathematical problem of constructing linear state-space representations of a stationary process has been studied in some depth in the past three decades see [5, 17, 1, 2, 57, 13, 36, 37, 39].

On the other hand, system identification, i.e. the *statistical* problem of describing an observed time series by a linear dynamic model, in particular by a state-space model of the type (1), has traditionally been regarded as a different problem. We shall argue in this paper that stochastic realization theory provides not only a rigorous and clear mathematical background for identification but that state-space identification can be seen, in a very specific sense, as essentially *the same problem* as stochastic realization.

Identification can be approached from (at least) two conceptually different viewpoints.

Identification by parametric optimization This is the traditional"optimization" approach, based on the principle of minimizing a suitable scalar measure of the discrepancy between the observed data and the data described by the probability law underlying a certain chosen model class. Well-known examples of distance functions are the *likelihood functions*, or the average squared *prediction-*

[1] Wide-sense stationarity will be simply referred to as stationarity hereafter. Also all random quantities encountered in this paper will be assumed to have zero mean.

error of the observed data corresponding to a particular model. Except for trivial model classes, these distance functions depend nonlinearly on the model parameters and the minimization can only be done numerically. Hence the optimization approach leads to iterative algorithms in the space of the parameters, say in the space of minimal (A, B, C, D) matrix quadruples which parametrize a chosen model class. In spite of the fact that this has been almost the only accepted paradigm in system Identification in the past three decades, [44, 60], this approach has several well-known drawbacks, among which the fact that the cost function generally has complicated local minima which, for moderate or large dimension of the model are very difficult to detect, there is often inherent insensitivity of the cost to variations of some parameters and corresponding ill-posedness of the estimation problem, there are difficulties to take consistently into account the (unknown) initial conditions, so that the methods only work "asymptotically", etc.

These limitations, it seems to us, are a consequence of the intrinsically "blind" philosophy which underlies setting the problem as a parameter optimization problem. For, almost all problems of control and/or estimator design could be (and in the past have sometimes been) formulated as parametric optimization problems after choosing a dynamic structure for the controller or the estimator. Pushing this philosophy to the extreme, in principle one would not need the maximum principle, Kalman filtering, H^∞ theory, etc. (in fact one would not need system theory altogether), since everything could be reduced to a nonlinear programming problem in a suitable space of controller or estimator parameters. It is very doubious however, whether any real progress in the field of control and estimation could have occurred by following this type of paradigm.

Identification by "subspace methods" i.e. stochastic realization This approach has more or less implicitely been suggested in the past by several authors [8, 13, 45] but is first clearly presented in [64] and is further analyzed and extended in [41, 52, 53]. The denomination "subspace method" is used in this paper specifically referring to the prototype procedure described below.

Subspace methods are based on the idea of constructing first a *state space* for the process y, starting from certain vector spaces, namely the *future* and *past* spaces (at a certain instant of time) associated to the observations. The state space is constructed by a geometrical operation on the data spaces, say by orthogonal projection of the future onto the past. Successively, a well conditioned basis is chosen in the state space e.g. by principal components (canonical correlation) analysis. Once a basis, i.e. a state vector, is chosen, the parameters A, C of the model are uniquely specified. The final step of finding the B and D matrices requires solving a Riccati equation.

The philosophy of constructing the state space as a first step of the model building procedure, is exactly the same underlying stochastic realization, the only difference being that the latter is a representation problem formulated in an abstract framework, where the future and past data spaces are Hilbert

spaces spanned by linear combinations of random variables of the process. The statistical problem of identification must be formulated instead in terms of the observed data. The basic mathematical structure of the two problems is however the same. In particular one recognizes by this similarity that the inherent nonlinearity of model identification has to do with the quadratic nature of the spectral factorization problem. It is well-known that spectral factorization for state-space models involves the solution of a Riccati equation (or more generally of a linear matrix inequality), a problem which has been object of intensive theoretical and numerical studies in the past three decades. The non linearity of the stochastic system identification problem is hence of a well-known and well understood kind and is much better dealt with by the explicit methods of Riccati solution developed in system theory rather than by non-specific optimization algorithms. In essence, the new paradigm for identification of state-space models is to use the procedures of geometric[2] stochastic realization theory suitably translated into algorithms of numerical linear algebra operating on the data.

As it will be discussed in Sections 2 and 6, the abstract Hilbert space operations of stochastic realization theory will have "concrete" counterparts in a Hilbert space $\mathbf{H}$ generated by shifted tail sequences of the observed data. In this setting the operations of stochastic realization may be regarded as statistical operations on the observed data. So we obtain a statistical model building theory which is perfectly isomorphic to the abstract probabilistic realization theory.

Classical Stochastic Realization

The $m \times m$ *spectral density* matrix of a purely-non-deterministic (p.n.d. hereafter) zero-mean stationary process y is the matrix function

$$\Phi(z) = \sum_{t=-\infty}^{\infty} \Lambda(t) z^{-t}$$

where

$$\Lambda(t) := E\{y(t+k)y(k)'\} = E\{y(t)y(0)'\}.$$

It is well-known that the spectral density matrix of a process admitting a state-space realization is a rational function of z. This fact follows easily by the classical Kintchine and Wiener formula for the spectrum of a filtered stationary process [30, 71].

PROPOSITION 1.1. *The transfer function $W(z) = C(zI - A)^{-1}B + D$ of any state space representation (1) of the stationary process y, is a* spectral factor *of Φ, i.e.*

[2] In this paper, as it will be explained in a moment, the adjective "gometric" is used with a predominant meaning of "coordinate-free". It has nothing to do with Differential Geometry.

$$W(z)W(1/z)' = \Phi(z). \tag{2}$$

Indeed it can be checked directly that, writing

$$\Phi(z) = \Phi_+(z) + \Phi_+(1/z)' \tag{3}$$

where $\Phi_+(z) = \frac{1}{2}\Lambda(0) + \Lambda(1)z^{-1} + \Lambda(2)z^{-2} + \cdots$ is the "causal" (i.e. analytic outside of the unit circle) component of $\Phi(z)$, one has

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}' + \frac{1}{2}\Lambda(0), \tag{4}$$

where

$$\Lambda(0) = CPC' + DD', \qquad \bar{C}' = APC' + BD', \tag{5}$$

$P = P'$ being a solution of the Lyapunov equation $P = APA' + BB'$. In other words, the spectrum of a process y described by a state-space model (1) is a rational function expressible in parametric form directly in terms of the parameters of the realization. The explicit computation of the spectrum is due to KALMAN AND ANDERSON [24, 6].

"Classical" stochastic realization theory [5, 6, 17, 18] was developed in the late sixties. It deals with the inverse problem of computing the parameters A, B, C, D of a state space realization starting from a suitable parametrization of the spectrum or covariance function of the process.

By the above proposition, this inverse problem is a parametric version of the *minimal spectral factorization problem* where one looks for rational spectral factors $W(z) = C(zI - A)^{-1}B + D$, of minimal degree, of a spectral density matrix $\Phi(z)$. Of course one assumes here that the process y is p.n.d. and has a spectral density which is a rational function of $z = e^{j\omega}$.

Spectral factorizability of rational matrix functions has been effectively characterized in the sixties by KALMAN, YAKUBOVICH AND POPOV [25, 73, 55] mainly in the context of stability theory. The first application of the Kalman-Yakubovich-Popov theory to factorization of spectral density functions and to stochastic realization (called "stationary covariance generation") is due to ANDERSON [5, 6].

The main result of the theory states that the minimal degree spectral factors of $\Phi(z)$, assumed given in the parametric form 4, are in one-to-one correspondence with the symmetric $n \times n$ matrices P solving the *Linear Matrix Inequality* (LMI)

$$M(P) := \begin{bmatrix} P - APA' & \bar{C}' - APC' \\ \bar{C} - CPA' & \Lambda(0) - CPC' \end{bmatrix} \geq 0 \tag{6}$$

in the following sense:

Corresponding to each solution $P = P'$ of (6), consider the full column rank matrix factor $\begin{bmatrix} B \\ D \end{bmatrix}$ of $M(P)$,

$$M(P) = \begin{bmatrix} B \\ D \end{bmatrix} [B'D'] \tag{7}$$

(this factor is unique modulo right multiplication by orthogonal matrices) and form the rational matrix

$$W(z) := C(zI - A)^{-1}B + D. \tag{8}$$

Then (8) is a minimal realization of a minimal analytic spectral factor of $\Phi(z)$ and all minimal factors can be obtained in this way.

Under some mild regularity conditions (6) is equivalent to an *Algebraic Riccati Inequality* (ARI) [5, 18]. These inequalities have been much studied both from a theoretical and a numerical viewpoint. They play an important role in many areas of system theory such as stability theory, dissipative systems and are central in H^∞ control and estimation theory. It seems to be much less appreciated in the scientific community that they play a very basic role in modeling of stationary random signals as well. Certain solutions of the LMI (or of the ARI) have special probabilistic properties and are related to Kalman-filter or "innovations-type" realizations. We shall refer the reader to the literature [72, 18, 39, 42] for a full discussion of these aspects of the problem.

Geometric Stochastic Realization

The classical "wide-sense" realization theory is purely distributional as it says nothing about representation of random quantities in a truly probabilistic sense (i.e. how to generate the random variables or the sample paths of a given process, not just its covariance function). This was implicitely pointed out by KALMAN already in [27]. In the last two decades a *geometric* or *coordinate-free* approach to stochastic modeling has been put forward in a series of papers by LINDQUIST, PICCI, RUCKEBUSCH ET AL. [35, 36, 37, 57, 58, 59] which aims at the representation of random processes in this more specific sense. This motivation is also present in the early papers by AKAIKE [1, 2].

A main point of the geometric approach is that a stochastic state-space system is defined in terms of the conditional independence relation between past and future of the signals involved. This relation is intrinsically *coordinate-free* and in the present setting involves only linear subspaces of a given ambient Hilbert space of random variables, typically made of linear functionals of the variables of the process y to be modeled (but in some situations other random data may be used to construct the model).

It is by now very well understood how to construct state spaces and build realizations in a "constructive" manner by a series of geometric operations involving subspaces of the background Hilbert space available to the modeler. It is one of the main goals of this paper to persuade the reader that these geometric procedures form the conceptual basis of "subspace" algorithms for identification and that their numerical implementation can be done directly and naturally via modern numerical linear algebra for subspace computations. The crucial first step to make this possible is a proper identification of the

Hilbert space of random data in which modeling takes place. This will be the goal of the next section.

It should be said that most results in the stochastic realization literature are about modeling of "time series" but in most practical situations there are exogenous "input" signals whose effect needs to be taken into account in the modeling phase. In this paper we shall discuss some generalizations of the basic ideas of realization theory which extend the geometric modeling theory to random processes in presence of inputs. Some preliminary results in this direction have been presented in [53] while a more complete discussion is to appear in the forthcoming paper [54].

Finally, it should be said that the idea of formulating state-space identification as a stochastic realization problem is not entirely new and has been present in the literature for some time. In particular Faurre [14, 15, 16] seems to be the first who systematically attempts to formulate identification as stochastic realization. His context is however still heavily coordinate-dependent. It is actually the geometric viewpoint and the vector space (actually the Hilbert space) characterization of the state space as a subspace of a certain data space that allow systematic introduction of numeric Linear Algebra and efficient computational tools to solve the problem.

2. The Hilbert space of a time series

In this section we shall build the basic Hilbert space structure in which the geometry of subspace identification will be defined.

For reasons of clarity and mathematical simplicity, we shall initially consider an idealized situation in which the observed data

$$\{\dots, u_{-1}, u_0, u_1, \dots, u_t, \dots\} \quad \{\dots, y_{-1}, y_0, y_1, \dots, y_t, \dots\} \tag{9}$$

with $u_t \in \mathbb{R}^p$, $y_t \in \mathbb{R}^m$, form an *infinitely long* time series originating in the remote past at $-\infty$. The finite data length situation will be discussed in Sections 6 and 7. The geometric approach of this article is based on the following basic "statistical" assumption on the data.

Assumption 2.1. *For $N \to \infty$ and for any $\tau \geq 0$, the time averages*

$$\frac{1}{N+1} \sum_{t=t_0}^{N+t_0} \begin{bmatrix} u_{t+\tau} \\ y_{t+\tau} \end{bmatrix} \begin{bmatrix} u_t \\ y_t \end{bmatrix}' \qquad \tau \geq 0 \tag{10}$$

converge and the limits are independent of the initial time t_0.

This assumption can be read as a kind of "statistical regularity" of the (future) data. It is of course unverifiable in practice as it says something about data which have not been observed yet. Some assumption of this sort about the mechanism generating future data seems however to be necessary to even formulate the identification problem.

In a continuous-time setting, functions admitting an "ergodic" limit of the sample correlation function 10, have been studied in depth by Wiener in his famous work on Generalized Harmonic Analysis [69, 70]. Although a systematic translation of the continuous-time results of Wiener into discrete-time seems not to be available in the literature, it is quite evident that a totally analogous set of results holds also for discrete- time signals. In particular it is rather easy to show, by adapting Wiener's proof for continuous time, that the limits of the time averages 10

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{t=t_0}^{N+t_0} \begin{bmatrix} u_{t+\tau} \\ y_{t+\tau} \end{bmatrix} \begin{bmatrix} u_t \\ y_t \end{bmatrix}' := \Lambda(\tau) = \begin{bmatrix} \Lambda_{uu}(\tau) & \Lambda_{uy}(\tau) \\ \Lambda_{yu}(\tau) & \Lambda_{yy}(\tau) \end{bmatrix} \quad \tau \geq 0, \quad (11)$$

form a matrix function Λ of *positive type*, in other words a *bona-fide stationary covariance matrix* sequence. We shall call Λ the *true covariance* of the time series $\{y_t, u_t\}$.

Now, for each $t \in \mathbb{Z}$ define the $p \times \infty$ and $m \times \infty$ matrices

$$u(t) := [u_t, u_{t+1}, u_{t+2}, \ldots] \quad (12a)$$
$$y(t) := [y_t, y_{t+1}, y_{t+2}, \ldots] \quad (12b)$$

and consider the sequences $u := \{u(t) \,|\, t \in \mathbb{Z}\}$ and $y := \{y(t) \,|\, t \in \mathbb{Z}\}$. These sequences will play a very similar role to two jointly stationary processes u and y, as referred to in the previous section.

Define the vector spaces $\mathcal{U}$ and $\mathcal{Y}$ of scalar semi-infinite real sequences obtained as finite linear combinations of the components of u and y,

$$\mathcal{U} := \{\sum a_k' u(t_k); \qquad a_k \in \mathbb{R}^p,\ t_k \in \mathbb{Z}\} \quad (13)$$
$$\mathcal{Y} := \{\sum a_k' y(t_k); \qquad a_k \in \mathbb{R}^m,\ t_k \in \mathbb{Z}\} \quad (14)$$

These vector spaces can be seen as the row spaces of two infinite block-Hankel matrices having as block rows the semi-infinite entries $u(t)$ and $y(t)$ of (12) where t is running over $\mathbb{Z}$.

NOTATIONS 1. In what follows the symbols $\vee$, $+$ and $\oplus$ will denote vector sum, *direct* vector sum and *orthogonal* vector sum of subspaces, the symbol $\mathbf{X}^\perp$ will denote the orthogonal complement of a (closed) subspace $\mathbf{X}$ of a Hilbert space with respect to some predefined ambient space. Given a collection $\{X_\alpha \,|\alpha \in A\}$ of subsets of a Hilbert space $\mathbf{H}$ we shall denote by $\overline{\text{span}}\{X_\alpha \,|\alpha \in A\}$ the closure in $\mathbf{H}$ of the linear (real) vector space generated by the collection. The orthogonal projection onto the subspace $\mathbf{X}$ will be denoted by the symbol $E(\cdot|\mathbf{X})$ or by the shorthand $E^{\mathbf{X}}$. The notation $E(z|\mathbf{X})$ will be used also when z is vector-valued. The symbol will then denote the vector with components $E(z_k|\mathbf{X})$, $k = 1, \ldots$. For vector quantities, $|v|$ will denote Euclidean length (or absolute value in the scalar case).

The vector sum $\mathcal{U} \vee \mathcal{Y}$ will originate our basic ambient space. This space can be naturally made into an inner product space in the following way.

First, define the bilinear form $\langle \cdot, \cdot \rangle$ on the generators of the space by letting

$$\begin{aligned} &\langle a' \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}, b' \begin{bmatrix} u(j) \\ y(j) \end{bmatrix} \rangle := \\ &\lim_{N\to\infty} \tfrac{1}{N+1} \sum_{t=0}^{N} a' \begin{bmatrix} u_{t+k} \\ y_{t+k} \end{bmatrix} \begin{bmatrix} u_{t+j} \\ y_{t+j} \end{bmatrix}' b = a' \Lambda(k-j) b, \end{aligned} \tag{15}$$

for $a,\ b \in \mathbb{R}^{p+m}$ (the prime denotes transpose), and then extend it by linearity to all elements of $\mathcal{U} \vee \mathcal{Y}$.

Let $\mathbf{a} := \{a_k,\ k \in \mathbb{Z}\}$ be a sequence of vectors $a_k \in \mathbb{R}^{p+m}$, with compact support in $\mathbb{Z}$, and let $\mathbf{a}' := \{a_k'\}$. A generic element ξ of the vector space $\mathcal{U} \vee \mathcal{Y}$ can be represented as

$$\xi = \sum_k a_k' \begin{bmatrix} u(k) \\ y(k) \end{bmatrix} := \mathbf{a}' \begin{bmatrix} u \\ y \end{bmatrix}$$

Introducing the infinite block-symmetric positive semidefinite Toeplitz matrix

$$T = \begin{bmatrix} \Lambda(0) & \Lambda(1) & \dots & \Lambda(k) & \dots \\ \Lambda(1)' & \Lambda(0) & \Lambda(1) & \dots & \dots \\ \vdots & & \ddots & & \vdots \\ \Lambda(k)' & & & \Lambda(0) & \\ \dots & & & & \end{bmatrix} \tag{16}$$

constructed from the "true" covariance sequence $\{\Lambda(0), \Lambda(1), \dots, \Lambda(k), \dots\}$ of the data, the bilinear form 15 on $\mathcal{U} \vee \mathcal{Y}$ can be represented by the quadratic form

$$\langle \xi, \eta \rangle = \langle \mathbf{a}' \begin{bmatrix} u \\ y \end{bmatrix}, \mathbf{b}' \begin{bmatrix} u \\ y \end{bmatrix} \rangle = \sum_{kj} a_k' \Lambda(k-j) b_j = \mathbf{a}' T \mathbf{b}.$$

We shall hereafter identify elements whose difference has norm zero (this means $\langle \xi, \xi \rangle = 0 \Leftrightarrow \xi = 0$). From the expression above it can be seen that the bilinear form is nondegenerate (unless $\Lambda = 0$ identically) and defines a bona-fide inner product. In the following we shall assume that for every k, the square block-Toeplitz matrix T_k in the upper left-hand corner of T, is positive definite. By closing the vector space $\mathcal{U} \vee \mathcal{Y}$ with respect to convergence in the norm induced by the inner product (15), one obtains a real Hilbert space $\mathbf{H} := \mathbf{U} \vee \mathbf{Y}$ (the wedge now means closed vector sum). This is the basic data space on which hereafter the models will be defined.

If in the limits of the sum (15) $t = 0$ is replaced by an arbitrary initial instant t_0 the limit does not change, so that

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

for all t_0. This is wide-sense stationarity in the present setting. The *shift operator* σ defined on the family of semi-infinite matrices (12), by setting

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

is then a linear map which is isometric with respect to the inner product (15) and can be extended by linearity to all of $\mathbf{H}$. The family $\{\sigma^t \,|\, t \in \mathbb{Z}\}$ with $\sigma^{-1} := \sigma^*$, forms a group of unitary operators on $\mathbf{H}$ which will be called the *shift*.

This Hilbert space framework for time series has been introduced in [41]. It is shown in this reference that the "stationary Hilbert space" $(\mathbf{H}, \sigma)$ is isomorphic to the standard stochastic Hilbert space setup widely used in the L^2-theory of second-order stationary random processes [31, 56, 47]. By virtue of this isomorphism one can formally think of the observed time series 9 as an ergodic sample path of some Gaussian stationary stochastic process $(\mathbf{u}, \mathbf{y})$ defined on a true probability space and having joint covariance matrices equal to the limit (11) of the sum (10) as $N \to \infty$.

Linear functions and operators on the tail sequences u and y correspond to the same linear functions and operators on the random variables of the processes $\mathbf{u}$ and $\mathbf{y}$. In particular the second order moments of the two random processes can be computed in terms of the tail sequences u and y, by substituting expectations with ergodic limits of the type (15). Since second order properties are all what matters in this paper, one may even regard the tail sequences u and y of (12) as being the *same* object as the two underlying stochastic process $\mathbf{u}$ and $\mathbf{y}$. The usual probabilistic language can be adopted in the present setting provided one identifies real random variables as semi-infinite strings of numbers having the "ergodic property" described at the beginning of this section. This will be done hereafter in the rest of this paper. The inner product of two semi-infinite strings ξ and η in $\mathbf{H}$ corresponds in particular to the expectation $E\{\xi\eta\}$. For reasons of uniformity of notatin we shall denote the inner product as expectation,

$$\langle \xi, \eta \rangle = E\{\xi\eta\}. \tag{17}$$

In the following $E\{\cdot\}$ will be allowed to operate on matrices, taking inner products row by row.

This unification of language permits to carry over in its entirety the geometric theory of stochastic realization derived in the abstract L^2 setting of [37, 39, 38] to the present framework. One may just reinterpret everything in the current setting, starting from the definition of $\mathbf{U}_t^-$, $\mathbf{Y}_t^-$, $\mathbf{U}_t^+$, $\mathbf{Y}_t^+$, the *past* and *future* subspaces of the "processes" u and y at time t. These are defined as the closure of the linear vector spaces spanned by the relative past or future "random variables" $u(t)$ and $y(t)$, in the metric of the Hilbert space $\mathbf{H}$. We shall use the notations

$$\begin{aligned} \mathbf{U}_t^- &:= \overline{\text{span}}\{\, u(s) \,|\, s < t \,\} \\ \mathbf{Y}_t^- &:= \overline{\text{span}}\{\, y(s) \,|\, s < t \,\} \end{aligned}$$

$$\begin{aligned} \mathbf{U}_t^+ &:= \overline{\text{span}}\{\, u(s) \mid s \geq t\} \\ \mathbf{Y}_t^+ &:= \overline{\text{span}}\{\, y(s) \mid s \geq t\} \end{aligned}$$

Note that, according to a widely accepted convention, the present is included in the future only and not in the past. The only difference to keep in mind here is the different interpretation that representation formulas like 1 have in the new context. The equalities involved in the representation

$$\begin{cases} x(t+1) &= Ax(t) + Bw(t) \\ y(t) &= Cx(t) + Dw(t) \end{cases} \tag{18}$$

are now to be understood in the sense of equalities of elements of $\mathbf{H}$, i.e. as asymptotic equality of sequences in the sense of Cesàro limits. In particular the equality signs in the model 18 do not necessarily imply that the same relations hold for the sample values y_t, x_t, w_t at a particular instant of time t. This is in a certain sense similar to the "with probability one" interpretation of the equality sign to be given to the model 18 in case the variable are bona- fide random variables in a probability space.

Modeling and estimation of stationary processes on infinite or semi- infinite time intervals, naturally involves various linear operations on the variables of the process which are *time-invariant*, i.e. independent of the particular instant of time chosen as a "present". In this setting it is possible (and convenient) to fix the present instant of time to an arbitrary value say $t = 0$ and work as if time was "frozen" at $t = 0$. At the occurrence one then "shifts" the operations in time by the action of the unitary operator σ^t on the data. In particular, the future and past subspaces of the processes y and u will often be referred at time $t = 0$ and denoted $\mathbf{Y}^+$ and $\mathbf{Y}^-$. For an arbitrary present instant t we have

$$\mathbf{Y}_t^+ = \sigma^t \mathbf{Y}^+, \quad \mathbf{Y}_t^- = \sigma^t \mathbf{Y}^-.$$

Consider the orthogonal projection $E[\,\xi\,|\mathbf{X}]$ of a (row) random variable ξ onto a subspace $\mathbf{X}$ of the space $\mathbf{H}$. In the probabilistic L^2 setting this has the well-known interpretation of wide- sense conditional expectation given the random variables in $\mathbf{X}$ (of a true conditional expectation, in the case of Gaussian distributions). In this setting the projection operator has an immediate and useful *statistical* meaning.

Assume for simplicity that $\mathbf{X}$ is given as the rowspace of some matrix of generators X, then the projection $E[\,\xi\,|\mathbf{X}]$ has exactly the familiar aspect of the least squares formula expressing the best approximation of the vector ξ as a linear combination of the rows of X. For, writing $E[\,\xi\,|\,X\,]$ to denote the projection expressed (perhaps nonuniquely) in terms of the rows of X, the classical linear "conditional expectation" formula leads to

$$E[\,\xi\,|\,X\,] = \xi X'[XX']^\sharp X, \tag{19}$$

which is the universally known "least squares" formula of statistics. The pseudoinverse $\sharp$ can be substituted by a true inverse in case the rows of X are linearly independent.

3. Input-Output models

It is known [7, 20, 21, 51] that identification of a causal input-output relation in the presence of feedback is an ill-posed problem (this is true of course in the absence of any specific information on the structure of the feedback link) and in this case the problem is better formulated as the identification of the *joint process* (y, u) on the basis of the joint corresponding observed time-series. This in turn falls into the general setup of time-series identification. In this paper we shall discuss specifically the case when there is *absence of feedback* from y to u, in the observed data. This concept will be formalized below.

Feedback-free processes

The appropriate setup for discussing feedback-free models is the theory of *feedback* and *causality* between stationary processes à la Granger [23]. See also [11, 7, 20, 21, 51]. We shall rephrase it in the language of conditionally orthogonal subspaces. The notation $\mathbf{A} \perp \mathbf{B}|\mathbf{X}$ means that the two subspaces $\mathbf{A}, \mathbf{B} \subset \mathbf{H}$ are *conditionally orthogonal* given a third subspace $\mathbf{X}$, i.e.

$$\langle \alpha - E^{\mathbf{X}}\alpha, \beta - E^{\mathbf{X}}\beta \rangle = 0 \quad \text{for} \quad \alpha \in \mathbf{A},\ \beta \in \mathbf{B}. \tag{20}$$

When $\mathbf{X} = 0$, this reduces to the usual orthogonality $\mathbf{A} \perp \mathbf{B}$. Conditional orthogonality is orthogonality after subtracting the orthogonal projections onto $\mathbf{X}$. This concept is discussed in depth in [37, 39].

One says that there is *absence of feedback* from y to u in the sense of Granger, 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. In our Hilbert space setup this is written as,

$$\mathbf{U}_t^+ \perp \mathbf{Y}_t^- \,|\, \mathbf{U}_t^- \tag{21}$$

where $\mathbf{U}_t^-$, $\mathbf{Y}_t^-$, $\mathbf{U}_t^+$, $\mathbf{Y}_t^+$ are the past and future subspaces of the processes u and y at time t.

This conditional orthogonality condition will be another basic assumption held throughout this paper. It is quite easy to see that, in conditions of absence of feedback, the "causal estimation error" process

$$y_s(t) := y(t) - E[y(t)|\mathbf{U}_{t+1}^-] \tag{22}$$

coincides with $y(t) - E[y(t)|\mathbf{U}]$ and hence is uncorrelated with the whole history of the input process u,

$$y_s(t) \perp \mathbf{U} \text{ for all } t \tag{23}$$

see [52]. Hence the process y_s may be called the *stochastic component of* y.

It also follows that the stochastic process y_u defined by the complementary projection

$$y_u(t) := E[y(t) \,|\, \mathbf{U}] \qquad t \in \mathbf{Z}$$

is uncorrelated with y_s. It will be named the *deterministic component of* y.

In the present feedback-free setting there is a natural, unique, "input- output" linear model

$$y(t) = y_u(t) + y_s(t) = E[\, y(t) \,|\, u(s); s \leq t \,] + y_s(t) \tag{24}$$

where $E[\, y(t) \,|\, u(s); s \leq t \,]$ is the best (in the sense of minimum variance of the error) estimate of the output $y(t)$ based on the past of u up to time t. Under some regularity conditions on the input process to be made precise later on, this estimate is described by a causal and stable linear convolution operator. Identifying the model (24) means identifying both the input-output "deterministic" part (described by a transfer function $W(z)$) and the additive "noise" process y_s. This last component is always present and may well be the most important for a realistic description of the output.

It is obvious that state-space descriptions for the process y can be obtained by combining two separate state-space models for y_s and y_u. For example, a (forward) innovation representation of y is obtained by combining together the (forward) innovation representation of y_s

$$\begin{aligned} x_s(t+1) &= A_s x_s(t) + B_s e_s(t) && \text{(25a)} \\ y_s(t) &= C_s x_s(t) + e_s(t) && \text{(25b)} \end{aligned}$$

where $e_s(t)$ is the one-step prediction error of the process y_s based on its own past, i.e. the (forward) innovation process of y_s, and the "deterministic" state-space model for y_u

$$\begin{aligned} x_u(t+1) &= A_u x_u(t) + B_u u(t) && \text{(26a)} \\ y_u(t) &= C_u x_u(t) + D_u u(t). && \text{(26b)} \end{aligned}$$

The process e_s has then the meaning of conditional innovation of y [52].

By combining together (25) and (26), the state-space innovation model of the process y "with inputs" has the following canonical structure,

$$\begin{aligned} \begin{bmatrix} x_s(t+1) \\ x_u(t+1) \end{bmatrix} &= \begin{bmatrix} A_s & 0 \\ 0 & A_u \end{bmatrix} \begin{bmatrix} x_s(t) \\ x_u(t) \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ B_u \end{bmatrix} u(t) + \begin{bmatrix} B_s \\ 0 \end{bmatrix} e_s(t) \\ y(t) &= \begin{bmatrix} C_s & C_u \end{bmatrix} \begin{bmatrix} x_s(t) \\ x_u(t) \end{bmatrix} + D_u u(t) + e_s(t) \end{aligned} \tag{27}$$

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 $G(z) = D_s + C_s(zI - A_s)^{-1}B_s$) often used in the identification literature.

It may happen that, even if the realizations of the two subsystems (stochastic and deterministic) are minimal, (27) may give redundant descriptions of the signals in certain particular cases, as there may be loss of observability when the transfer functions $W(z)$ and $G(z)$ have common poles and common corresponding eigenspaces.

These cases are highly non-generic and in practice one need not worry about this unlikely event in black-box identification. However, in certain structured problems there may be some a priori 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.

4. Constructing the State space of the "stochastic" component

The theme of this section will be a review of geometric realization theory for the stochastic component of y. Since there will be no input processes in this section, for notational simplicity the subscript "s" will be dropped.

The geometric theory centers on the idea of *Markovian Splitting Subspaces* for the process y. This concept is the probabilistic analog of the deterministic notion of state space of a dynamical system and captures at an abstract level the property of "dynamic memory" that the state variables have in deterministic system theory. Once a stochastic state space is given the construction of the auxiliary random quantities which enter in the model and in particular the state process is fairly obvious. The state vector $x(t)$ of a particular realization can be regarded just as a particular basis for the state space, hence once a state-space is constructed, finding state equations is just a matter of choosing a basis and computing coordinates.

Let y be a stationary vector process and $\mathbf{Y}$ the relative Hilbert space of linear functionals. Let $\mathbf{X}$ be a subspace of some large stationary Hilbert space $\mathbf{H}$ of wide-sense random variables containing $\mathbf{Y}$. Define

$$\mathbf{X}_t := \sigma^t \mathbf{X}, \quad \mathbf{X}_t^- := \vee_{s \le t} \mathbf{X}_s, \quad \mathbf{X}_t^+ := \vee_{s \ge t} \mathbf{X}_s.$$

Definition 4.1. *A* Markovian Splitting Subspace $\mathbf{X}$ *for the process y is a subspace of* $\mathbf{H}$ *making the vector sums* $\mathbf{Y}^- \vee \mathbf{X}^-$ *and* $\mathbf{Y}^+ \vee \mathbf{X}^+$ *conditionally orthogonal (i.e. uncorrelated) given* $\mathbf{X}$*, denoted,*

$$\mathbf{Y}^- \vee \mathbf{X}^- \perp \mathbf{Y}^+ \vee \mathbf{X}^+ \,|\, \mathbf{X}. \tag{28}$$

The conditional orthogonality condition 28 can be equivalently written as

$$E[\mathbf{Y}^+ \vee \mathbf{X}^+ \,|\, \mathbf{Y}^- \vee \mathbf{X}^-] = E[\mathbf{Y}^+ \vee \mathbf{X}^+ \,|\, \mathbf{X}] \tag{29}$$

which gives the intuitive meaning of the splitting subspace $\mathbf{X}$ as a dynamic memory of the past for the purpose of predicting the joint future.

The subspace $\mathbf{X}$ is called *proper*, or *purely-non- deterministic* if

$$\cap_t \mathbf{Y}_t^- \vee \mathbf{X}_t^- = \{0\}, \text{ and } \cap_t \mathbf{Y}_t^+ \vee \mathbf{X}_t^+ = \{0\}.$$

Obviously for the existence of proper splitting subspaces y must also be purely non deterministic [56]. Properness is, by the Wold decomposition theorem, equivalent to the existence of two vector white noise processes w and $\bar{w}$ such that,

$$\mathbf{Y}^- \vee \mathbf{X}^- = \mathbf{H}^-(w), \quad \mathbf{Y}^+ \vee \mathbf{X}^+ = \mathbf{H}^+(\bar{w})$$

Here the symbols $\mathbf{H}^-(w)$, $\mathbf{H}^+(w)$ etc. denote the Hilbert subspaces linearly generated by the past and future of the process w. The spaces

$$\mathbf{S} := \mathbf{Y}^- \vee \mathbf{X}^- \quad \text{and} \quad \overline{\mathbf{S}} := \mathbf{Y}^+ \vee \mathbf{X}^+ \tag{30}$$

associated to a Markovian Splitting subspace $\mathbf{X}$, play an important role in the geometric theory of stochastic systems. They are called the *scattering pair* of $\mathbf{X}$ as they can be seen to form an incoming-outgoing pair in the sense of Lax-Phillips Scattering Theory [34].

DEFINITION 4.2. *Given a stationary Hilbert space* $(\mathbf{H}, \sigma)$ *containing* $\mathbf{Y}$, *a* scattering pair *for the process* y *is a pair of subspaces* $(\mathbf{S}, \overline{\mathbf{S}})$ *satisfying the following conditions,*

1. $\sigma^*\mathbf{S} \subset \mathbf{S}$ *and* $\sigma\overline{\mathbf{S}} \subset \overline{\mathbf{S}}$, *i.e.* $\mathbf{S}$ *and* $\overline{\mathbf{S}}$ *are invariant for the left and right shift semigroups (this means that* $\mathbf{S}_t$ *is increasing and* $\overline{\mathbf{S}}_t$ *is decreasing with time).*
2. $\mathbf{S} \vee \overline{\mathbf{S}} = \mathbf{H}$
3. $\mathbf{S} \supset \mathbf{Y}^-$ *and* $\overline{\mathbf{S}} \supset \mathbf{Y}^+$
4. $\mathbf{S}^\perp \subset \overline{\mathbf{S}}$ *or, equivalently,* $\overline{\mathbf{S}}^\perp \subset \mathbf{S}$

The following representation Theorem provides the link between Markovian splitting subspaces and scattering pairs.

THEOREM 4.3 ([37]) *The intersection*

$$\mathbf{X} = \mathbf{S} \cap \overline{\mathbf{S}} \tag{31}$$

of any scattering pair of subspaces of $\mathbf{H}$ *is a Markovian splitting subspace. Conversely every Markovian splitting subspace can be represented as the intersection of a scattering pair. The correspondence* $\mathbf{X} \leftrightarrow (\mathbf{S}, \overline{\mathbf{S}})$ *is one-to-one, the scattering pair corresponding to* $\mathbf{X}$ *being given by*

$$\mathbf{S} = \mathbf{Y}^- \vee \mathbf{X}^- \qquad \overline{\mathbf{S}} = \mathbf{Y}^+ \vee \mathbf{X}^+. \tag{32}$$

The process of forming scattering pairs associated to $\mathbf{X}$ should be thought of as an "extension" of the past and future spaces of y. The rationale for this extension is that scattering pairs have an extremely simple splitting geometry due to the fact that

$$\mathbf{S} \perp \overline{\mathbf{S}} \,|\, \mathbf{S} \cap \overline{\mathbf{S}} \tag{33}$$

which is called *perpendicular intersection*. It is easy to show that Property 4. in the definition of a scattering pair is actually equivalent to perpendicular intersection. This property of conditional orthogonality given the intersection can also be seen as a natural generalization of the Markov property[3]. Note

[3] In which case $\mathbf{S} = \mathbf{X}^-$, $\overline{\mathbf{S}} = \mathbf{X}^+$ and $\mathbf{X} = \mathbf{X}^- \cap \mathbf{X}^+$.

that $\mathbf{A} \perp \mathbf{B} \,|\, \mathbf{X} \Rightarrow \mathbf{A} \cap \mathbf{B} \subset \mathbf{X}$ but the inclusion of the intersection in the splitting subspace $\mathbf{X}$ is only *proper* in general. For perpendicularly intersecting subspaces, the intersection is actually the *unique minimal subspace* making them conditionally orthogonal.

Any basis vector $x(0) := [x_1(0), x_2(0), \ldots, x_n(0)]'$ in a (finite- dimensional) Markovian splitting subspace $\mathbf{X}$ generates a stationary Markov process $x(t) := \sigma^t x(0), t \in \mathbb{Z}$ which serves as a *state* of the the process y. If $\mathbf{X}$ is proper, the Markov process is purely non deterministic.

Denote by $\mathbf{W}_t$, $\bar{\mathbf{W}}_t$ the spaces spanned by the components, at time t, of the generating noises $w(t)$ and $\bar{w}(t)$, of the scattering pair of $\mathbf{X}$. Since

$$\mathbf{S}_{t+1} = \mathbf{S}_t \oplus \mathbf{W}_t, \tag{34}$$

we can write,

$$\mathbf{X}_{t+1} = \mathbf{S}_{t+1} \cap \overline{\mathbf{S}}_{t+1} = (\mathbf{S}_t \cap \overline{\mathbf{S}}_{t+1}) \oplus (\mathbf{W}_t \cap \overline{\mathbf{S}}_{t+1}) \tag{35}$$

Since $\overline{\mathbf{S}}_t$ is decreasing in time, we have $\mathbf{S}_t \cap \overline{\mathbf{S}}_{t+1} \subset \mathbf{X}_t$ and by projecting the shifted basis $\sigma x(t) := x(t+1)$, onto the last orthogonal direct sum above, the time evolution of any basis vector $x(t) := [x_1(t), x_2(t), \ldots, x_n(t)]'$ in $\mathbf{X}_t$ can be represented by a linear equation of the type $x(t+1) = Ax(t) + Bw(t)$. It is also easy to see that by the p.n.d. property, A must have all its eigenvalues strictly inside of the unit circle. Naturally, by decomposing instead $\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{S}}_t \oplus \overline{\mathbf{W}}_t$ we could have obtained a *backward difference equation* model for the Markov process x, driven by the backward generator $\bar{w}$.

To complete the representation, note also that by definition of the past space, $y(t) \in (\mathbf{S}_{t+1} \cap \overline{\mathbf{S}}_t)$. Inserting the decomposition 34 and projecting $y(t)$ leads to a state-output equation of the form $y(t) = Cx(t) + Dw(t)$. Here one could also obtain a state-output equation driven by the backward noise $\bar{w}$, the same noise driving the backward state model obtained before.

As we have just seen, any basis in a Markovian splitting subspaces produces a stochastic realization of y. It is easy to reverse the implication. In fact the following fundamental characterization holds.

THEOREM 4.4. *[35, 39] The state space* $\mathbf{X} = \text{span}\{ x_1(0), x_2(0), \ldots, x_n(0)\}$ *of any stochastic realization (1) is a* Markovian Splitting Subspace *for the process* y. *Conversely, given a finite-dimensional Markovian splitting subspace* $\mathbf{X}$, *to any choice of basis* $x(0) = [\, x_1(0), x_2(0), \ldots, x_n(0)]'$ *in* $\mathbf{X}$ *there corresponds a stochastic realization of* y *of the type (1).*

Once a basis in $\mathbf{X}$ is available, there are obvious formulas expressing the coefficient matrices A, C and $\bar{C}$ in terms of the processes x and y.

$$A = Ex(t+1)x(t)'\,P^{-1} \tag{36}$$

$$C = Ey(t)x(t)'\,P^{-1} \tag{37}$$

$$\bar{C} = Ey(t-1)x(t)' \tag{38}$$

where P is the Gramian matrix of the basis (equal to the state covariance matrix). The matrices B and D however are related to the (unobservable) generating white noise w and require the solution of the LMI.

Stochastic realizations are called *internal* when $\mathbf{H} = \mathbf{Y}$, i.e. the state space is built from the Hilbert space made just of the linear statistics of the process y. For identification the only realizations of interest are the internal ones.

A central problem of geometric realization theory is to construct and to classify the *minimal* state spaces, i.e. the minimal Markovian splitting subspaces for the process y.

The obvious ordering of subspaces of $\mathbf{H}$ by inclusion, induces an ordering on the family of Markovian splitting subspaces. The notion of minimality is most naturally defined with respect to this ordering. Note that this definition is independent of assumptions of finite-dimensionality and applies also to infinite-dimensional Markovian splitting subspaces, i.e. to situations where comparing dimension would not make much sense.

DEFINITION 4.5. *A Markovian splitting subspace is* minimal *if it does not contain (properly) other Markovian splitting subspaces.*

The study of minimality forms an elegant chapter of stochastic system theory. There are several known geometric and algebraic characterizations of minimality of splitting subspaces and of the corresponding stochastic state- space realizations. Since however the discussion of this topic would take us too far from the main theme of the paper we shall refer the reader to the literature [37, 39].

Contrary to the deterministic situation minimal Markovian splitting subspaces are *non unique*. Two very important examples are the *forward and backward predictor spaces* (at time zero):

$$\mathbf{X}_- := \mathrm{E}^{\mathbf{H}^-}\mathbf{H}^+ \quad \mathbf{X}_+ := \mathrm{E}^{\mathbf{H}^+}\mathbf{H}^- \tag{39}$$

for which we have the following characterization [37].

PROPOSITION 4.1. *The subspaces $\mathbf{X}_-$ and $\mathbf{X}_+$ are the unique minimal splitting subspaces contained in the past $\mathbf{H}^-$, and, respectively, in the future $\mathbf{H}^+$, of the process y.*

A basis in the forward predictor space $\mathbf{X}_-$ originates a stationary state-space model in which the state variables are linear functionals of the past history of the process y, i.e. $x(t) \in \mathbf{Y}_t^-$. In other words the state coincides with its best estimate (the orthogonal projection), $E[x(t) \mid \mathbf{Y}_t^-]$ given the past of y. It follows that the dynamical equations 1 describe in this case a steady-state Kalman predictor and the input white noise $w = w_-$ is the steady state *innovation process* of y.

5. Stationary realization of the deterministic component

In order to construct the state-space of realizations of the deterministic component y_u which are driven by the (not necessarily white) process u, it is necessary to generalize the geometric theory of stochastic realization of the previous section. In order to streamline notations, we shall here too, drop the subscript u and whenever possible fix $t = 0$.

Assume that y and u are two jointly stationary p.n.d. processes of dimensions m and p. We shall call a model of the type

$$x(t+1) = Ax(t) + Bu(t) \tag{40a}$$

$$y(t) = Cx(t) + Du(t) \tag{40b}$$

a *deterministic realization of* y with input process u. Models of this kind reduce to the standard (Markovian) stochastic models when, 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 $y(t)$ can be expressed as functionals of the infinite past (and present) of u, i.e. $x(t) \in \mathbf{U}_t^-$ and $y(t) \in \mathbf{U}_{t+1}^-$. Realizations with this property will be called *causal*.

The following technical assumption of "sufficient richness" of the input process will facilitate the geometric constructions of this section.

Assumption 5.1. *For each t the input space $\mathbf{U}$ admits the direct sum decomposition*

$$\mathbf{U} = \mathbf{U}_t^- + \mathbf{U}_t^+ \tag{41}$$

An analogous condition (namely $\mathbf{U}_t^- \cap \mathbf{U}_t^+ = 0$) is discussed in [41] 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 [56] Chapter II, Sect. 7.

It will also be assumed all through this section that $y(t) \in \mathbf{U}_{t+1}^-$ (this is the feedback-free property). Because of this property and in virtue of assumption 5.1, $y(t)$ has a unique representation as a causal functional

$$y(t) = \sum_{-\infty}^{t} W_{t-k}\, u(k). \tag{42}$$

where $\hat{W}(z) = \sum_0^{+\infty} W_k z^{-k}$ is analytic in $\{|z| > 1\}$. Indeed, $\hat{W}(z)$ is just the transfer function of the Wiener filter $y(t) = E[y(t) | \mathbf{U}_{t+1}^-]$ and can be expressed as

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

where $G(z)$ is the outer (or minimum-phase) spectral factor of Φ_u and the symbol $[\cdot]_+$ means "analytic part", see e.g. [56] Chapter II. It is evident that $\hat{W}(z)$ is analytic and, because of nonsingularity of Φ_u on the unit circle, unique almost everywhere. Hence the feedback-free assumption implies that the input-output map relating u to y must be a causal map. It follows that a minimal state-space realizations of y must necessarily be causal. Our interest here will henceforth be on causal realizations.

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

DEFINITION 5.2. *We shall call a subspace* $\mathbf{X} \subset \mathbf{U}^-$ *a* (causal) oblique splitting subspace *for the pair* $(\mathbf{Y}^+, \mathbf{U}^-)$ *if*

$$E_{||\mathbf{U}^+}[\mathbf{Y}^+ \vee \mathbf{X}^+ \,|\, \mathbf{U}^-] = E_{||\mathbf{U}^+}[\mathbf{Y}^+ \vee \mathbf{X}^+ \,|\, \mathbf{X}]. \tag{43}$$

Note that this condition is a generalization of the conditional orthogonality condition 29 of the Markovian case. For the reason explained a moment ago, we shall only consider causal splitting subspaces in this paper so the "incoming" subspace $\mathbf{S}$ will always be fixed equal to $\mathbf{U}^-$. The *oblique predictor space* $\mathbf{X}^{+/-} := E_{||\mathbf{U}^+}[\mathbf{Y}^+ \,|\, \mathbf{U}^-]$ is an example of oblique splitting subspace which is obviously contained in $\mathbf{U}^-$.

Write

$$y(t) = (H_W u)(t) + (W^+ u)(t) \tag{44}$$

where,

$$(H_W u)(t) := \sum_{-\infty}^{-1} W_{t-k} u(k), \quad (W^+ u)(t) := \sum_{0}^{t} W_{t-k} u(k). \tag{45}$$

Evidently $(H_W u)(t) \in \mathbf{U}^-$ and $(W^+ u)(t) \in \mathbf{U}^+$ for $t \geq 0$. For $t \geq 0$ the random variable $(H_W u)(t)$ is the oblique projection of $y(t)$ onto $\mathbf{U}^-$ and hence $\mathbf{X}^{+/-} = \overline{\text{span}}\{\, (H_W u)(t) \,|\, t \geq 0\}$. It is easy to see that $\mathbf{X}^{+/-}$ is contained in all causal oblique splitting subspaces $\mathbf{X} \subset \mathbf{U}^-$, or, equivalently said, is the (unique) *minimal* causal oblique splitting subspace.

The following results are "deterministic" counterparts of the characterizations of Markovian splitting subspaces stated in the previous section. Proofs can be found in [53, 54].

PROPOSITION 5.1. *Let* $\mathbf{X}$ *be an oblique splitting subspace and define the corresponding "extended future space"*

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

Then

$$\mathbf{X} = E_{||\mathbf{U}^+}[\bar{\mathbf{S}} \,|\, \mathbf{U}^-] = E_{||\mathbf{U}^-}[\mathbf{U}^- \,|\, \bar{\mathbf{S}}] = \bar{\mathbf{S}} \cap \mathbf{U}^-.$$

(This may be read as "oblique intersection" of $\bar{\mathbf{S}}$ *and* $\mathbf{U}^-$ *).*

This representation result in particular applies to the extended future space $\bar{\mathbf{Y}}^+ = \mathbf{Y}^+ \vee (\mathbf{X}^{+/-})^+$ (which is in a sense the "minimal" possible extended future space $\bar{\mathbf{S}}$).

THEOREM 5.3. *The oblique predictor space can be computed as the intersection*

$$\mathbf{X}^{+/-} = (\mathbf{Y}^+ \vee \mathbf{U}^+) \cap \mathbf{U}^-. \tag{46}$$

Note that it is in general *not true* that $\mathbf{X}^{+/-} = \mathbf{Y}^+ \cap \mathbf{U}^-$ as the inclusion

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

is *proper*, unless some special condition is satisfied. The reader should be warned of the fact that some statements in the paper [45] may be seen as an endorsement of this erroneous intersection representation.

PROPOSITION 5.2. *Let the symbols have the same meaning as in Proposition 5.1. Then*

$$\bar{\mathbf{S}} = (\bar{\mathbf{S}} \cap \mathbf{U}^-) + (\bar{\mathbf{S}} \cap \mathbf{U}^+). \tag{47}$$

This intersection representation extends the formula $\bar{\mathbf{S}} = (\bar{\mathbf{S}} \cap \mathbf{S}) \oplus (\bar{\mathbf{S}} \cap \mathbf{S}^\perp)$, known for "orthogonal" splitting subspaces [37, 39].

The following argument, based on the decomposition 47 shows how state space realizations can be constructed by a procedure based on the geometry of oblique splitting subspaces. The reader may appreciate the similarity of the reasoning with that used in the "stochastic" construction of Section 4.

Denote by $\mathbf{U}_t$ the p-dimensional subspace of $\mathbf{U}_t^+$ spanned by the components of $u(t)$. By Assumption 5.1

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

and by the decomposition 47 we can write

$$\bar{\mathbf{S}}_{t+1} \cap \mathbf{U}_{t+1}^- = (\bar{\mathbf{S}}_{t+1} \cap \mathbf{U}_t^-) + (\bar{\mathbf{S}}_{t+1} \cap \mathbf{U}_t). \tag{48}$$

Now pick a basis vector $x(t)$, say of dimension[4] n in $\mathbf{X}_t$ and let $x(t+1)$ be the corresponding vector shifted by one unit of time. The n scalar components of $x(t+1)$ span $\bar{\mathbf{S}}_{t+1} \cap \mathbf{U}_{t+1}^-$ and, since $\bar{\mathbf{S}}_{t+1} \subset \bar{\mathbf{S}}_t$, we have

$$(\bar{\mathbf{S}}_{t+1} \cap \mathbf{U}_t^-) \subset \mathbf{X}_t$$

so, by projecting $x(t+1)$ onto the two components of the direct sum decomposition (48) we obtain a unique representation of the type

$$x(t+1) = Ax(t) + Bu(t).$$

[4] Here for the sake of illustration we assume that $\mathbf{X}_t$ is finite-dimensional.

Similarly, since $y(t) \in \mathbf{U}_{t+1}^-$, we have

$$y(t) \in \bar{\mathbf{S}}_t \cap \mathbf{U}_{t+1}^- = (\bar{\mathbf{S}}_t \cap \mathbf{U}_t^-) + (\bar{\mathbf{S}}_t \cap \mathbf{U}_t)$$

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

$$y(t) = Cx(t) + Du(t).$$

The following Theorem is the analog of Theorem 4.4.

THEOREM 5.4. *The state space of any causal realization of y with input process u is an oblique splitting subspace.*

Conversely, assume the joint spectral density of y and u is rational and that the input process satisfies Assumption 5.1. Then the oblique predictor subspace $\mathbf{X}^{+/-}$ is finite dimensional. To any choice of a basis vector $x(t)$ in a finite-dimensional oblique splitting subspace $\mathbf{X}_t$, there correspond unique matrices (A, B, C, D) such that the representation

$$x(t+1) = Ax(t) + Bu(t) \tag{49a}$$

$$y(t) = Cx(t) + Du(t), \tag{49b}$$

holds and the realization 49 is causal, i.e. $|\lambda(A)| < 1$.

Computations of the system matrices by oblique projections appear profusely in recent papers dealing with the identification of "deterministic" systems [68, 67, 65]. The theory of state space realization exposed in this section explains the reason why this should in a sense be the natural type of computation in this type of problems. See also [29, 54].

6. Finite-interval Realization with inputs

The analysis in the previous Sections 3, 4 and 5 is based on the idealized assumption that one has access to a doubly infinite sequence of data. In real experiments only a *finite* string of observed data

$$\{u_0, u_1, \ldots, u_t, \ldots, u_N\} \quad \{y_0, y_1, \ldots, y_t, \ldots, y_N\} \tag{50}$$

is available, where, however, N may often be quite large. More specifically, we shall assume that N is sufficiently large that replacing the ergodic limits 11 by the finite time averages of $N+1$ elements, yields good approximations of the true covariances

$$\{\Lambda(0), \Lambda(1), \Lambda(2) \ldots, \Lambda(T)\}, \tag{51}$$

for some lag $T << N$. This is equivalent to saying that N is sufficiently large for the time averages 10 to be essentially the same as the inner products

$$E\{\begin{bmatrix} u(\tau) \\ y(\tau) \end{bmatrix} \begin{bmatrix} u(0) \\ y(0) \end{bmatrix}'\} = \Lambda(\tau)$$

for $\tau = 0, \ldots, T$. Hence, for the purpose of theoretical analysis of the algorithms, we may proceed as if we had two finite sequences of random vectors

$$\{u(0), u(1), u(2), \ldots, u(T)\}, \quad \{y(0), y(1), y(2), \ldots, y(T)\}, \tag{52}$$

where each $u(t)$ and $y(t)$ is still a semi-infinite string of data of the type 12. It is of course to be understood that, when it comes to practical implementation, the strnigs will be finite and some approximations will have to be made. For example the inner products will have to be replaced by finite time averages. These approximations are unavoidable and an analysis of their effects requires a priori assumptions on the data, for example of the probabilistic type as it is made in classical statistics. This is somewhat outside the scope of this paper and will not be discussed further here.

Now, the state-space construction of Sections 4 and 5 was done in a stationary setting, where the state-space model has to represent the output process on an *infinite* time horizon. In a situation where only the finite segment of data 52 is available it is necessary to understand the relation between the data and models which realize them on a *finite* interval of time. Even if the ultimate goal of modeling and identification is the construction of a *stationary model* describing the data, when only a finite segment of data is available it is important to view model building as the construction of finite-interval realizations. The reason is that this viewpoint only gives the correct way of dealing with the unknown initial conditions that have unavoidably to be attached to a stationary model on a finite interval.

A finite-interval realization describes the process on a finite interval $[0, T]$ without bringing in the history of the process outside of $[0, T]$ (i.e. no unknown initial conditions). However, even if the process is stationary, a finite-interval realization turns out to be in general a time- varying (nonstationary) system. The notion of Markovian splitting subspace applies without difficulty to finite-interval realizations [35, 9, 41]. For example, it is easy to see that the *finite-interval predictor spaces* for the (stochastic component of the) process y

$$\hat{\mathbf{X}}_t^{+/-} := E^{\mathbf{Y}_{[0,t)}} \mathbf{Y}_{[t,T]} \quad \hat{\mathbf{X}}_t^{-/+} := E^{\mathbf{Y}_{[t,T]}} \mathbf{Y}_{[0,t)} \tag{53}$$

are minimal Markovian splitting subspaces for the process y on the finite interval $[0, T]$. Chosing a basis $\hat{x}(t)$ in the forward predictor space $\hat{\mathbf{X}}_t^{+/-}$ leads to a state space realization which is a "transient Kalman filter" (innovation) representation of the process of the type

$$\hat{x}(t+1) = A\hat{x}(t) + B(t)\epsilon(t) \quad \hat{x}(0) = 0, \tag{54}$$

$$y(t) = C\hat{x}(t) + D(t)\epsilon(t) \tag{55}$$

where the state $\hat{x}(t)$ can be interpreted as the orthogonal projection onto $\mathbf{Y}_t^-$ of the state $x(t)$ of any minimal stationary realization of y in a suitable basis. Note that the Kalman filter realization has the same (A, C) pair of the stationary model, while $B(t) := K(t)R(t)^{1/2}$, $D(t) := R(t)^{1/2}$ are time-varying, determined by the solution of a Riccati equation. The process $\epsilon(t) := R(t)^{-1/2}[y(t) - C\hat{x}(t)]$ is a normalized m-dimensional white noise process: the *normalized transient innovation process* of y on $\{t \geq 0\}$.

Doally, choosing a *dual* basis $\bar{x}(t)$ in the backward predictor space $\hat{\mathbf{X}}_t^{-/+}$ leads to a state space realization which is a *backward* "transient Kalman filter" (or a backward innovation) representation of the process. This finite-interval realization involves naturally a backward recursion for the state, i.e.

$$\begin{array}{rcll} \bar{x}(t) & = & A'\bar{x}(t+1) + \bar{B}(t)\bar{\epsilon}(t+1) & \bar{x}(T) = 0, \\ y(t) & = & \bar{C}\bar{x}(t) + \bar{D}(t)\bar{\epsilon}(t) & \end{array} \tag{56}$$

on $\{t \leq T\}$, where $\bar{B}(t) := \bar{K}(t)\bar{R}(t)^{1/2}$, $\bar{D}(t) := \bar{R}(t)^{1/2}$ are determined by a backward Riccati equation and $\bar{\epsilon}(t) := -\bar{R}(t)^{-1/2}[y(t) - \bar{C}\bar{x}(t)]$ is a normalized m-dimensional white noise process: the *normalized backward transient innovation process* of y on $\{t \leq T\}$.

Note that these state space models are initialized at $\hat{x}(0) = 0$ and to $\bar{x}(T) = 0$ respectively.

Using state-space models here facilitates things greatly as the transient Kalman filter realizations above have exactly the same constant parameters (A, C) and $\bar{C}$ as the stationary model. Hence the $(A, C, \bar{C})$ parameters of a minimal stationary realization of the process y can be computed *exactly* from the finite data 52 by the formulas

$$A = E\hat{x}(t+1)\hat{x}(t)'\, P(t)^{-1} \tag{57}$$

$$C = Ey(t)\hat{x}(t)'\, P(t)^{-1} \tag{58}$$

$$\bar{C} = Ey(t-1)\hat{x}(t)'. \tag{59}$$

Here $P(t)$ is the Grammian (covariance) matrix of the basis $\hat{x}(t)$. *Note that this computation does not require a preliminary estimation of the initial conditions.*

One may compare with the fact that the parameters $(A, C, \bar{C})$ of a minimal (deterministic) realization 4 of the spectral density (or of the covariance function $\Lambda(\tau)$) of y can be computed exactly from a finite sequence of covariance estimates 51, see [43].

The above lies at the grounds of the success of Van Overschee and De Moor "subspace" approach to time-series identification [64]. For a more detailed analysis of the method one may consult [41, 43].

Now while subspace-methods identification of "purely stochastic" systems (i.e. of signals or time series) seems to be reasonably well-understood, for systems which are driven by "inputs" or, *exogenous* variables, the picture is far less clear. The various algorithms given in the literature are not based on clear realization principles [46, 68, 65] and some require a rather complicated analysis to motivate [65]. In particular the finite-interval modeling issue seems to have been largely overlooked. The theoretical grounds of stochastic realization in presence of inputs have been clarified only recently [52, 53, 54].

Regarding this last point, it has been shown in [52] that when data are finite, the two (stochastic and deterministic) realization problems cannot be decoupled and solved separately as it can instead be done for the stationary infinite-interval situation. This is an important point to keep in mind, especially for what concerns the order estimation step of the identification algorithm which

may perhaps sound surprising to practitioners used to neglect initial conditions and only "think asymptotically".

The finite history subspaces of $\mathbf{H}$ generated by the finite stochastic data 52 will be denoted

$$\begin{aligned} \mathbf{U}_{[0,T]} &:= \text{span}\{\, u(t) \mid 0 \le t \le T \,\} \\ \mathbf{Y}_{[0,T]} &:= \text{span}\{\, y(t) \mid 0 \le t \le T \,\} \end{aligned}$$

The orthogonal complement of $\mathbf{U}_{[0,T]}$ in $\mathbf{U}_{[0,T]} \vee \mathbf{Y}_{[0,T]}$ will be denoted by $\mathbf{U}^{\perp}_{[0,T]}$, so that $\mathbf{U}_{[0,T]} \oplus \mathbf{U}^{\perp}_{[0,T]} = \mathbf{U}_{[0,T]} \vee \mathbf{Y}_{[0,T]}$. The practical computation of $\mathbf{U}^{\perp}_{[0,T]}$ will be addressed in Section 7.

The following statement is straightforward.

LEMMA 6.1. *Let y_u be described by the deterministic realization (26). Then the process*

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

is described by the same state-space model as y_u but started at a different initial state, namely

$$\begin{aligned} \hat{x}_u(t+1) &= A_u \hat{x}_u(t) + B_u u(t) && \text{(60a)} \\ \hat{y}_u(t) &= C_u \hat{x}_u(t) + D_u u(t) && \text{(60b)} \\ \hat{x}_u(0) &= E[x_u(0) \mid \mathbf{U}_{[0,T]}]. && \text{(60c)} \end{aligned}$$

It follows from this Lemma that the deterministic part of (27), namely the system matrices (A_u, B_u, C_u, D_u) (in a suitable basis) can be identified by using a "deterministic" identification procedure, based on the data $\{\, \hat{y}_u(t),\ u(t) \mid t = 1, 2, \ldots, T\}$. Once the system matrices (of a minimal realization) are computed the estimate of the initial state $\hat{x}_u(0)$ can also be reconstructed.

The identification of the stochastic subsystem can be based on the projections of the output data onto the complementary subspace $\mathbf{U}^{\perp}_{[0,T]}$. To this end, we introduce the random vectors

$$\hat{y}_s(t) := y(t) - E[y(t) \mid \mathbf{U}_{[0,T]}] \quad 0 \le t \le T$$

These can be computed from the available data and actually we have

$$\mathbf{U}^{\perp}_{[0,T]} := \text{span}\{\, \hat{y}_s(t) \mid 0 \le t \le T \,\}$$

The following Proposition, taken from [52], shows how the finite-time estimate $\hat{y}_s(t)$ relates to the 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*

$$\hat{y}_s(t) = y_s(t) + \tilde{y}_u(t), \qquad 0 \le t \le T. \tag{61}$$

Hence for finite data length, the projection $\hat{y}_s(t)$ of the output on the complementary subspace $\mathbf{U}^{\perp}_{[0,T]}$ 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 in the estimate of the initial state of the deterministic component, $\tilde{x}_u(0) := x_u(0) - \hat{x}_u(0)$. In fact,

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

This additional term is a source of difficulty in identification of the stochastic part since if it is not properly subtracted off from the $\hat{y}_s$'s, it tends to produce a stochastic model of y_s of a much higher dimension than the true order $\tilde{n}$. In fact, the estimated model will tend to include also the dynamics of the deterministic subsystem. Therefore a preliminary step is necessary for the identification of the stochastic component, i.e. to filter out $\tilde{y}_u(t)$ somehow.

7. A GENERAL "SUBSPACE" IDENTIFICATION ALGORITHM

In this section we shall briefly examine some algorithmic aspects of the finite interval realization procedure. We shall fix a "present" time $t = k$ at half way between the extremes of the interval $[0, T]$ (actually to fix notations we shall just assume that $T = 2k - 1$) and organize our (finite) data in "past" and "future" block Hankel matrices of the form

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

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

and

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

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

The relative rowspaces, denoted $\mathbf{Y}_k^-, \mathbf{U}_k^-, \mathbf{Y}_k^-, \mathbf{U}_k^+$ are the "past" and "future" spaces of the data at time k. In practice with data of finite length, the inner product in the Hilbert space $\mathbf{H}$ is approximated by the Euclidean inner product

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

and this makes the stochastic realization procedures of the previous sections very simple to implement as vector space computations in the Euclidean space $\emptyset R^N$.

The practical computation of $\mathbf{U}_{[0,T]}^{\perp}$ can be done by an LQ factorization of the data matrix $\begin{bmatrix} U \\ Y \end{bmatrix}$ generating $\mathbf{U}_{[0,T]} \vee \mathbf{Y}_{[0,T]}$. (For simplicity of notations we shall not indicate the finite interval endpoint explicitely in the full data matrices.)

The LQ factorization has the form

$$\begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} L_{uu} & 0 \\ L_{yu} & L_{yy} \end{bmatrix} \begin{bmatrix} Q_u' \\ Q_y' \end{bmatrix} \tag{62}$$

where $Q_u' Q_u = I$, $Q_y' Q_y = I$, $Q_u' Q_y = 0$ and L_{uu}, L_{yy} are lower triangular. This gives immediately the block Hankel matrices of the *finite-interval* deterministic and stochastic components of y as

$$\hat{Y}_u = E[Y|\mathbf{U}_{[0,T]}] = Y Q_u Q_u' = L_{yu} Q_u'$$

and

$$\hat{Y}_s = E[Y|\mathbf{U}_{[0,T]}^{\perp}] = Y Q_y Q_y' = L_{yy} Q_y'.$$

The formulas follow immediately by noting that the rows of Q_u' form an orthonormal basis for the rowspace of U and those of Q_y' an orthonormal basis for the orthogonal complement $\mathbf{U}_{[0,T]}^{\perp}$ in $\mathbf{U}_{[0,T]} \vee \mathbf{Y}_{[0,T]}$.

In the basis constituted by the rows of U, we have $E[Y|U] = L_{yu} L_{uu}^{-1} U$. If the data were infinitely long, one could check causality from u to y (equivalently absence of feedback from y to u)by checking whether the matrix L_{yu} is lower triangular. This is however not necesarily true for data of finite lenght, for the reasons explained in the previous section.

For convenience, we also write the factorization (62) as

$$\begin{bmatrix} U_k^- \\ U_k^+ \\ Y_k^- \\ Y_k^+ \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 & 0 \\ L_{21} & L_{22} & 0 & 0 \\ L_{31} & L_{32} & L_{33} & 0 \\ L_{41} & L_{42} & L_{43} & L_{44} \end{bmatrix} \begin{bmatrix} Q_1' \\ Q_2' \\ Q_3' \\ Q_4' \end{bmatrix} \tag{63}$$

where the Q_i's are orthogonal matrices.

Identification of the Deterministic Part

Subspace methods to identify a stationary state space model for the deterministic component y_u, may be based on computing the state $\hat{x}_u(t)$ of the system 60 at two time instants, say k and $k+1$ and then solve in the least squares sense,

$$\begin{bmatrix} \hat{x}_u(k+1) \\ \hat{y}_u(k) \end{bmatrix} = \begin{bmatrix} A_u & B_u \\ C_u & D_u \end{bmatrix} \begin{bmatrix} \hat{x}_u(k) \\ u(k) \end{bmatrix}, \tag{64}$$

for the coefficient matrices A_u, B_u, C_u, D_u.

The theory of Section 5 works for finite-interval realization as well. Here we only have the finite history $\mathbf{U}_k^-$ available but the state $\hat{x}_u(k)$ of a minimal causal realization of $\hat{y}_u$ must still be a basis for the finite interval oblique predictor space at time k. This subspace in turn can be computed as the intersection of the extended future $\bar{\mathbf{Y}}_k^+ = \mathbf{Y}_k^+ \vee \mathbf{U}_k^+$ with the past space of u at time k. Below we give an algorithm to compute a well- conditioned basis in the intersection $\bar{\mathbf{Y}}_k^+ \cap \mathbf{U}_k^-$, based on the GSVD [48, 64].

Note that in order to get the right constant parameters in the realization, the basis at time $k+1$ should be chosen so as to correspond to the stationarily time-shifted state $E[x_u(k+1)|\,\mathbf{U}\,]$, $\quad x_u(k+1) = \sigma x_u(k)$, otherwise time varying matrices A_u, B_u, C_u, D_u are obtained. This means that after applying the basis selection algorithm at time $k+1$, a suitable linear transformation should be applied to make $\hat{x}_u(k+1)$ coherent with the basis chosen at the previous instant k.

Algorithm for computing a well conditioned basis for the subspace $\hat{\mathbf{X}} = \bar{\mathbf{Y}}^+ \cap \mathbf{U}^-$, based on the GSVD.

The basis can be chosen by using a variant of the subspace intersection algorithm described in [22]. This procedure also numerically determines the dimension $\hat{n}$ of $\hat{\mathbf{X}}$.

Observe that from (63) we get the LQ decomposition

$$\begin{bmatrix} U^- \\ \bar{Y}^+ \end{bmatrix} := \begin{bmatrix} U^- \\ U^+ \\ Y^+ \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ \bar{L}_{21} & \bar{L}_{22} \end{bmatrix} \begin{bmatrix} Q_1' \\ \bar{Q}_2' \end{bmatrix}$$

where $\bar{Q}_2'$ is obtained by stacking the rows of Q_2' over those of Q_4'. We have deleted the subscript k. We shall assume that $[\bar{L}_{21} \;\; \bar{L}_{22}]$ has full row rank. From the theory of Generalized Singular Valued Decomposition, ([48]), there exist orthogonal matrices V, Z and a nonsingular X of appropriate dimensions, such that

$$\begin{aligned} \bar{L}_{21} &= X\,C\,V' = X\,\mathrm{diag}(c_1, \cdots, c_\nu)V', \qquad 1 \geq c_1 \geq \cdots \geq c_\nu \geq 0 \\ \bar{L}_{22} &= X\,S\,Z' = X\mathrm{diag}(s_1, \cdots, s_\nu)Z', \qquad 1 \geq s_\nu \geq \cdots \geq s_1 \geq 0 \end{aligned}$$

where $CC' + SS' = I$ and $\nu \leq pk$. Setting

$$Q' := CV'Q_1' + SZ'\bar{Q}_2'$$

one has $\bar{Y}^+ = XQ'$ and $Q'Q = I$, so that the rows of Q_1' and $\hat{Q}'$ form orthogonal bases for the rowspaces $\mathbf{U}^-$ and $\bar{\mathbf{Y}}^+$, respectively. Moreover, we have $c_i = \cos(\theta_i)$, $i = 1, \cdots, \nu$, where θ_i is the i-th smallest principal angle between the rowspaces $\bar{\mathbf{Y}}^+$ and $\mathbf{U}^-$.

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 $V'Q_1$ span the intersection $\bar{\mathbf{Y}}^+ \cap \mathbf{U}^-$ [22]. Since this does rarely happen in actual computations, we must pick $\hat{n}$ in such a way that the first $\hat{n}$ principal angles are numerically zero.

Preprocessing of $\hat{\mathbf{y}}_s$

Note that (61) can be written more compactly in vector form as

$$\hat{\mathbf{y}}_s = \Omega_u \tilde{x}_u(0) + \mathbf{y}_s \tag{65}$$

where the boldface characters denote stacked vectors and Ω_u is the observability matrix of the deterministic system. The two terms on the right hand side are uncorrelated and once the parameters (A_u, C_u) of the deterministic subsystem have been identified, the matrix Ω_u can be assumed to be known.

It seems that one may easily filter out $\tilde{\mathbf{y}}_u$ from $\hat{\mathbf{y}}_s$ by just pre-multiplying the vector equation 65 by $\Omega_u \Omega_u^\perp$ where $\Omega_u^\perp$ is a matrix with rows spanning the left nullspace of Ω_u. With a procedure of this kind however a distortion is introduced on the time series $\{y_s(t),\ t = 0, \ldots, T\}$ which seems to be very hard to remove. To keep control on the reconstruction errors, one needs to filter out the unwanted term $\tilde{y}_u$ by a *sequential* algorithm. Below we shall describe an algorithm [53] which in principle only distorts a small finite initial segment of the time series $\{y_s(t),\ t = 0, \ldots, T\}$.

ALGORITHM 7.1. *Assume A_u and A_s have no common eigenvalues. The following algorithm recovers the time series $\{y_s(t),\ t = 0, \ldots, T\}$ exactly, up to the first ν sample values.*

1. *Compute a left-coprime factorization of the rational matrix $C_u(I - z^{-1}A_u)^{-1}$. Let the $m \times m$ and $m \times n$ polynomial matrices in the unit backward shift z^{-1}, $D(z^{-1}) = \sum_0^\nu D_k z^{-k}$, and $N(z^{-1}) = \sum_0^{\nu-1} N_k z^{-k}$ be such a left-coprime factorization, i.e. let*

$$D(z^{-1})^{-1} N(z^{-1}) = C_u(I - z^{-1}A_u)^{-1} \tag{66}$$

2. *Compute*

$$\check{y}(t) := D(z^{-1})\hat{y}_s(t) \tag{67}$$

so that $\check{y}(t) = N(z^{-1})\tilde{x}_u(0) + D(z^{-1})y_s(t)$. Since the first term has finite support $t = 0, \ldots, \nu - 1$,

$$\check{y}(t) = D(z^{-1})y_s(t) \quad t \geq \nu.$$

3. *Solve the following vector difference equation in the unknown variable* $z(t)$

$$D(z^{-1})z(t) = \check{y}(t) \qquad t \geq \nu \tag{68}$$

started with initial conditions $z(0) = z(1) = \ldots z(\nu - 1) = 0$. *Then* $z(t) = y_s(t), t \geq \nu$.

Identification of the Stochastic Part

The identification of the stochastic subsystem (25) is done by processing the "stochastic" data matrix $\hat{Y}_s$ computed in the previous step and can be done by the stochastic techniques described e.g in [8], [13]. A reliable method for computing the state-space realization in stochastically balanced form is presented in [64]. The algorithms of [64] are further streamlined in [43, 41]. As explained in this last reference, identification of the model in a *stochastically balanced* canonical form is to be recommended in order to guarantee positivity after truncations by retaining only the larger singular values (see Section 8 for a discussion of the positivity problem. The state (basis) vector in the predictor space should then be chosen to be the corresponding truncated subvector of canonical variates.

8. Concluding remarks

There are questions about the statistical significance (what are the uncertainty bounds on the parameters and on the estimated transfer functions etc.) of the subspace/realization approach to identification which are often asked.

In order to address these questions we shall recall a result about subspace identification given in [41, 43]. Identification by geometric stochastic realization methods has there been shown to be *equivalent* to the following three-step procedure,

1. The first step is estimation of the finite sequence of covariance matrices 51 from the observed data.

2. The second step is identification of a rational model for the covariance sequence. This is a minimal partial realization (also called "rational extension") problem. Given a finite set of "experimental " covariance data

$$\{\Lambda_0, \Lambda_1, \ldots, \Lambda_T\} \tag{69}$$

one is asked to find a minimal value of n and a minimal[5] triplet of matrices $(A, C, \bar{C})$, of dimensions $n \times n$, $m \times n$ and $m \times n$ respectively, such that

$$\Lambda_k = CA^{k-1}\bar{C}' \qquad k = 1, \ldots, T. \tag{70}$$

[5] Recall that $(A, C, \bar{C})$ is minimal if (A, C) is completely observable and $(A, \bar{C}')$ is completely reachable.

The solution of the partial realization problem leads to "estimates" of the parameters $(A, C, \bar{C})$ of a minimal realization of a rational spectral density matrix of the process.

3. The third step is to compute a stationary state-space model (typically the forward innovation model) by solving the Linear Matrix Inequality 6, or an appropriate equivalent Riccati equation, relative to the rational estimated spectrum computed in step two.

The main difference with the mainstream statistical approach is that the estimation of $(A, C, \bar{C})$ is not done *directly* by optimizing a likelihood or other distance functions but by just *matching second order moments* i.e. by solving the equations 70.

This way of proceeding is an instance of *estimation by the method of moments* described in the statistical textbooks e.g. [12, p. 497], a very old idea used extensively by K. Pearson in the beginning of the century. The underlying estimation principle is that the parameter estimates should match exactly the sample second order moments and is close in spirit to the wide-sense setting that we are working in. It does not involve optimality or minimal distance criteria between the "true" and the model distributions.

Are these estimates consistent, efficient etc.? One should note that the sample estimates of the second order moments Λ_k may very well be "maximum likelihood" or optimal in some sense, depending on which parent distribution one imagines to have generated the data. Very generally the estimates $\hat{\Lambda}_k$ can be chosen at least consistent (i.e. tending to the true second order moments as the sample size goes to infinity). In this case the method gives automatically *consistent* estimates of the *true partial covariance sequence* in the sense that at least T true moments $\Lambda_k \quad k = 1, 2, \ldots, T$ will be described exactly by the model as $N \to \infty$.

Some argue that estimation by the method of moments is in general "non-efficient" and it is generally claimed in the literature that one should expect better results (in the sense of smaller asymptotic variance of the estimates) by optimization methods. Actually if the covariance estimates are maximum likelihood estimates and the partial realization problem has a unique solution (modulo similarity), then, choosing $(A, C, \bar{C})$ in a canonical form, there is a continuous map

$$\{\Lambda_0, \Lambda_1, \ldots, \Lambda_T\} \to (A, C, \bar{C})$$

which is even locally one-to-one. It follows by a well- known Theorem of Zehna [74] that the canonical parameter estimates are also maximum likelihood. Hence in this case we have efficient estimates.

Positivity

A warning is in order concerning the implementation of the subspace identification methods described above in that some nontrivial mathematical questions related to positivity of the estimated spectrum have been completely overlooked

in the discussion. This issue is thoroughly discussed in [41] and here we shall just give a short summary. The problem occurs only in identification of the stochastic component, here named y for short.

In determining a minimal triplet $(A, C, \bar{C})$ interpolating the partial sequence (69) so that $CA^{k-1}\bar{C}' = \Lambda_k \quad k = 1, 2, \ldots, T$, we also completely determine the infinite sequence

$$\{\Lambda_0, \Lambda_1, \Lambda_2, \Lambda_3, \ldots\} \tag{71}$$

by setting $\Lambda_k = CA^{k-1}\bar{C}'$ for $k = T+1, T+2, \ldots$. This sequence is called a *minimal rational extension* of the finite sequence (69). The attribute "rational" is due to the fact that

$$Z(z) := \frac{1}{2}\Lambda_0 + \Lambda_1 z^{-1} + \Lambda_2 z^{-2} + \ldots = \frac{1}{2}\Lambda_0 + C(zI - A)^{-1}\bar{C}' \tag{72}$$

is a rational function. In order for (71) to be a bona fide covariance sequence, however, it is necessary, but *not* sufficient, that the Toeplitz matrix

$$T = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \Lambda_2 & \cdots & \Lambda_\nu \\ \Lambda_1' & \Lambda_0 & \Lambda_1 & \cdots & \Lambda_{\nu_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Lambda_\nu' & \Lambda_{\nu-1}' & \Lambda_{\nu-2} & \ldots & \Lambda_0 \end{bmatrix} \tag{73}$$

be nonnegative definite. In fact, it is required that the function (the spectral density corresponding to (71))

$$\Phi(z) = \Lambda_0 + \sum_{k=1}^{\infty} \Lambda_k (z^k + z^{-k}) = Z(z) + Z(z^{-1})' \tag{74}$$

be nonnegative on the unit circle. This is equivalent to the function $Z(z)$ being *positive real*. Consequently, the partial realization needs to be done subject to the extra constraint of positivity.

The constraint of positivity is a rather tricky one and in all identification methods which are directly or indirectly, as the subspace methods described in the literature, based on the interpolation condition (70) it is normally disregarded. For this reason these methods may fail to provide a positive extension and hence may lead to data $(A, C, \bar{C})$ for which there are no solutions of the LMI and hence to totally inconsistent results.

It is important to appreciate the fact that the problem of positivity of the extension has little to do with the "noise" or "sample variability" of the covariance data and is present equally well for a finite covariance sequence extracted from a true (infinitely long) rational covariance sequence. For there is no guarantee that, even in this idealized situation, the order of a minimal rational extension 71 of the finite covariance subsequence would be sufficiently high to equal the order of the infinite sequence and hence to generate a positive extension. A minimal partial realization may well fail to be positive because its order is too low to guarantee positivity.

Neglecting the positivity constraint amounts to tacitly assuming that

ASSUMPTION 8.1. *The finite covariance data (69) can be generated exactly by some (unknown) stochastic system whose dimension is equal to the rank of the block Hankel matrix*

$$H_\mu = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \cdots & \Lambda_\mu \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \cdots & \Lambda_{\mu+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Lambda_\mu & \Lambda_{\mu+1} & \Lambda_{\mu+2} & \cdots & \Lambda_{2\mu-1} \end{bmatrix}, \tag{75}$$

where $\mu = [\frac{T}{2}]$.

This assumption is not "generically satisfied" and it can be shown [41] that there are relatively "large" sets of data (69) for which it does not hold. It is not even enough to assume that the data is generated from a "true" finite-dimensional stochastic system: the rank condition is also necessary. Otherwise, for a minimal triplet $(A, C, \bar{C})$ which satisfies the interpolation condition (70), the positivity condition will not be automatically fulfilled, and the matrix A may even fail to be stable [10].

REFERENCES

1. H. AKAIKE (1975). Markovian representation of stochastic processes by canonical variables, *SIAM J. Control*, **13** , pp. 162–173.
2. H. AKAIKE (1974). Stochastic Theory of Minimal Realization. *IEEE Trans. Automatic Control*, vol. AC-19, no. 6, pp. 667–674.
3. H. AKAIKE (1976). Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. *System Identification: Advances and Case Studies*, R. MEHRA AND D. LAINIOTIS, EDS., Academic, pp. 27–96.
4. B.D.O. ANDERSON (1967). A System Theory Criterion for Positive- Real Matrices, *SIAM Journal on Control*, **5**, n.2, pp. 171–182.
5. B.D.O. ANDERSON (1969). The inverse problem of Stationary Covariance Generation, *Journal of Statistical Physics*, **1**, n.1, pp. 133–147.
6. B.D.O. ANDERSON (1967). An algebraic solution to the spectral factorization problem, *IEEE Trans. on Automatic Control* **AC-12** pp. 410–414.
7. B.D.O. ANDERSON AND M.R. GEVERS (1982). Identifiability of linear stochastic systems operating under linear feedback, *Automatica*, **18**, pp. 195–213.
8. M. AOKI (1990). *State Space Modeling of Time Series* (2nd ed.), Springer.
9. F. BADAWI, A. LINDQUIST AND M. PAVON (1979). A stochastic realization approach to the smoothing problem, *IEEE Trans. on Automatic Control*, **AC-24**, pp. 878–888.
10. C. I. BYRNES AND A. LINDQUIST (1982). The stability and instability of partial realizations, *Systems and Control Letters*, **2**, pp. 2301–2312.
11. P.E. CAINES AND C.W. CHAN (1976). Estimation, Identification and feedback, in *System Identification: Advances and Case Studies*, R. MEHRA AND D. LAINIOTIS, EDS., Academic Press.

12. H. Cramer, (1949). *Mathematical Methods of Statistics*, Princeton.
13. U.B. Desai, D. Pal and R.D. Kirkpatrick (1985). A Realization Approach to Stochastic Model Reduction. *Int. J. Control* **42**, pp. 821–838.
14. P. Faurre (1969). *Identification par minimisation d'une réprésentation Markovienne de processus aléatoires*, Symposium on Optimization, Nice , *Springer Lect. Notes in Math.*, **132**.
15. P. Faurre, P. Chataigner (1971). *Identification en temp reel et en temp differee par factorisation de matrices de Hankel*, French-Swedish colloquium on process control, IRIA Roquencourt.
16. P. Faurre and J.P. Marmorat (1969). Un algorithme de réalisation stochastique. *C. R. Academie Sciences Paris*, A, **268**, pp. 978–981.
17. P. Faurre (1973). *Realisations Markovienne de processus stationnaires*, Report de Recherche n. 13, INRIA.
18. P. Faurre, M. Clerget, and F. Germain (1979). *Opérateurs Rationnels Positifs*, Dunod.
19. P.A. Fuhrmann (1981). *Linear Operators and Systems in Hilbert Space*, McGraw Hill.
20. M.R. Gevers and B.D.O. Anderson (1981). Representation of jointly stationary feedback free processes, *Intern. Journal of Control* **33**, pp. 777–809.
21. M.R. Gevers and B.D.O. Anderson (1982). On jointly stationary feedback-free stochastic processes, *IEEE Trans. Automatic Control* **AC-27**, pp. 431–436.
22. G.H. Golub and C.R. Van Loan (1989). *Matrix Computations* (2nd ed.), The Johns Hopkins Univ. Press.
23. C.W.J. Granger (1963). Economic processes involving feedback, *Information and Control*, **6**, pp. 28–48.
24. R.E. Kalman (1963). On a new Criterion of Linear Passive Systems, *Proc. of the First Allerton Conference* Univ. of Illinois, pp. 456–470.
25. R.E. Kalman (1963). Lyapunov Functions for the problem of Luré in Automatic Control, *Proc. Natnl. Acad of Sciences (USA)*, **49**, pp. 201–205.
26. R.E. Kalman (1981). Realization of covariance sequences, *Proc. Toeplitz Memorial Conference* , Tel Aviv, Israel, Birkhäuser.
27. R.E. Kalman (1965). Linear stochastic filtering: reappraisal and outlook, *Proc. Symposium on Syatem Theory*, Politechnic Institute of Brooklin, pp. 197–205.
28. R.E. Kalman, P.L. Falb, and M.A. Arbib (1969). *Topics in Mathematical Systems Theory*, McGraw-Hill.
29. T. Katayama and G. Picci (1997). An approach to Realization of Stochastic Systems with exogenous inputs, submitted to *SYSID 97*, Kobe Japan.
30. A. Kintchine (1934). Korrelationstheorie Stationäre Stochastischen Prozessen, *Math. Annalen*, **109**, pp. 604–615.
31. (1941). Stationary Sequences in Hilbert space, *Bull. Math. Moscow State University*, **2**, p. 40.

32. W.E. Larimore (1990). System identification, reduced-order filtering and modeling via canonical variate analysis, *Proc. American Control Conference*, pp. 445–451.
33. W.E. Larimore (1990). Canonical Variate Analysis in Identification, Filtering, and Adaptive Control. *Proc. 29th IEEE Conference on Decision and Control*, pp. 596–604.
34. P.D. Lax and R.S. Phillips (1967). *Scattering Theory*, Academic Press, New York.
35. A. Lindquist, G. Picci and G. Ruckebusch (1979). On minimal splitting subspaces and Markovian representation, *Math. System Theory*, **12**: pp. 271–279.
36. A. Lindquist and G. Picci (1979). On the stochastic realization problem *SIAM J. Control and Optimization*, **17**, pp. 365–389.
37. A. Lindquist and G. Picci (1985). Realization theory for multivariate stationary Gaussian processes, *SIAM J. Contr. & Optimiz.* **23**, pp. 809–857.
38. A. Lindquist and M. Pavon (1984). On the structure of state space models of discrete-time vector processes, *IEEE Trans. on Automatic Control*, **AC-29**, pp. 418–432.
39. A. Lindquist and G. Picci (1991). A geometric approach to modelling and estimation of linear stochastic systems, *Journal of Mathematical Systems, Estimation and Control*, **1**, pp. 241–333.
40. A. Lindquist and G. Picci (1994). On "subspace-methods" identification and stochastic model reduction, *Proc. 10th IFAC Symposium on System Identification*, Copenhagen, DK, **2**, pp. 397–403.
41. A. Lindquist and G. Picci (1996). Canonical correlation analysis approximate covariance extension and identification of stationary time series, *Automatica*, **32**, pp. 709–733.
42. A. Lindquist, G. Michaletzky and G. Picci (1995). Zeros of Spectral Factors, the Geometry of Splitting Subspaces, and the Algebraic Riccati Inequality, *SIAM J. Control & Optimization*, **33**, pp. 365–401.
43. A. Lindquist and G. Picci (1996). Geometric Methods for State Space Identification, in *Identification, Adaptation, Learning*, (Lectures given at the NATO-ASI School, *From Identifiation to Learning* held in Como, Italy, Aug.1994), Springer Verlag.
44. L. Ljung (1987). *System Identification – Theory for the User*. Prentice-Hall.
45. M. Moonen, B. De Moor, L. Vanderberghe and J. Vandewalle (1989). On- and Off-Line Identification of Linear State-Space Models. *Int. J. Control* **49**, pp. 219–232.
46. M. Moonen and J. Vandewalle (1990). QSVD Approach to On- and Off- Line State-Space Identification. *Int. J. Control* **51**, pp. 1133–1146.
47. J. Neveu (1968). *Processus aleatoires Gausssiens*, Les Presses de l'Universitè de Montreal, Montreal.
48. C.C. Paige and M.A. Saunders (1981). Towards a generalized Singular Value Decomposition, *SIAM Journal on Numerical Analysis*, **18**, pp. 398–

405.
49. G. Picci (1976). Stochastic realization of Gaussian processes *Proc. of the IEEE*, **64**, pp. 112–122.
50. G. Picci (1991). Stochastic Modeling and Stochastic Realization Theory, in *Mathematical System Theory: the influence of R.E. Kalman* A. Antoulas ed., pp. 213–229, Springer Verlag.
51. G. Picci and S. Pinzoni (1986). Dynamic factor-analysis models for stationary processes. *IMA Journal on Mathematics of Control and Information* Vol.**3**, No. 2 , pp. 185–210.
52. G. Picci and T. Katayama (1996). Stochastic realization with exogenous inputs and "Subspace Methods" Identification, *Signal Processing*, (in press).
53. G. Picci and T. Katayama (1996). "A simple "subspace" identification algorithm with exogenous inputs", *Proceedings of the 1996 triennial IFAC Congress*, vol **I**, pp. 175–181.
54. G. Picci. Stochastic Realization and System Identification (submitted for publication.
55. V.M. Popov (1964). Hyperstability and Optimality of Automatic Systems with several Control functions, *Revue Rumaine des Sciences Techniques, Ser. Elechtrotechnique et Energetique*, **9**, pp. 629–690.
56. N.I. Rozanov (1963). *Stationary Random Processes.* Holden-Day.
57. G. Ruckebusch (1976). Répresentations markoviennes de processus gaussiens stationnaires, *C.R.Acad.Sc.Paris* , Series A, **282**, pp. 649–651.
58. G. Ruckebusch (1978). A state space approach to the stochastic realization problem, *Proc. 1978 IEEE Intern. Symp. Circuits and Systems*, pp. 972–977.
59. G. Ruckebusch (1978). Factorisations minimales de densités spectrales et répresentations markoviennes, *Proc. 1re Colloque AFCET–SMF*, Palaiseau, France.
60. T. Söderström and P. Stoica (1989). *System Identification.* Prentice-Hall.
61. J.H. van Schuppen (1994). Stochastic realization of a Gaussian stochastic control system, *Acta Applicandae Methematicae* **35** , pp. 193–212.
62. J.H. van Schuppen (1989). Stochastic realization problems, in *Three decades of mathematical system theory*, H. Nijmeijer and J.M. Schumacher eds., *Lecture notes in Control and Information Sciences*, **135**, pp. 480–532, Springer Verlag.
63. G. Picci and J.H. van Schuppen (1992). Stochastic realization of finite–valued processes and Primes in the Positive Matrices, in *Recent advances in Mathematical Theory of Systems, Control,Networks and Signal Processing II*, H. Kimura, S. Kodama eds. (Proceedings of MTNS-91, Kobe 1991), pp. 227–232, Mita press, Tokyo.
64. P. Van Overschee and B. De Moor (1993). Subspace algorithms for the stochastic identification problem, *Automatica* **29**, pp. 649–660.
65. P. Van Overschee and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic– stochastic systems,

Automatica, **30**, pp. 75–93.
66. P. Van Overschee and B. De Moor (1994). A unifying theorem for subspace system identification algorithms and its interpretation. *Proc. 10th IFAC Symposium on System Identification* **2**, pp. 145–156.
67. M. Verhaegen and P. DeWilde (1993). Subspace model identification. Part I: The output-error state-space identification class of algorithms, *International Journal of Control*, **56**, pp. 1187–1210, (1992); Part II: analysis of the elementary output-error state-space model identification algorithm, *International Journal of Control*, **56**, pp. 1211–1241, (1992): Part III: analysis of the ordinary output-error state-space model identification algorithm, *International Journal of Control*, **58**, pp. 555–586.
68. M. Verhaegen (1994). Identification of the deterministic part of MIMO State Space Models given in Innovations form from Input-Output data, *Automatica* **30**, pp. 61–74.
69. N. Wiener (1930). Generalized Harmonic Analysis. *Acta Mathematica*, **55**, pp. 117–258.
70. N. Wiener (1933). Generalized Harmonic Analysis, in *The Fourier Integral and Certain of its Applications*, Cambridge U.P.
71. N. Wiener (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, The M.I.T. Press, Cambridge, Mass.
72. J.C. Willems (1971). Least Squares Stationary Optimal Control and the Algebraic Riccati Equation, *IEEE Trans. Automatic Control*, **AC-16**, pp. 621–634.
73. V.A. Yakubovich (1963). Absolute stability of Nonlinear Control Systems in critical case: part I and II, *Avtomatika i Telemekhanika*, **24**, pp. 293–303, 717–731.
74. P.W. Zehna (1966). Invariance of maximum likelihood estimation, *Annals of Math. Statistics*, **37**, p. 755.
75. H.P. Zeiger and A.J. McEwen (1974). Approximate linear realization of given dimension via Ho's algorithm, *IEEE Trans. Automatic Control*, vol. **AC-19**, p. 153.