Recursive subspace identification based on instrumental variable unconstrained quadratic optimization[*]

G. Mercère, S. Lecoeuche and M. Lovera [†]

Abstract

The problem of the recursive formulation of the MOESP class of subspace identification algorithms is considered and two novel instrumental variable approaches are introduced. The first one leads to an RLS-like implementation, the second to a gradient type iteration. The relative merits of both approaches are analysed and discussed, while simulation results are used to compare their performance with the one of existing techniques.

1   Introduction

Recursive subspace model identification (RSMI) has been an active area of research for the last ten years [1, 2, 3, 4, 5]. In particular, the application of these techniques has been recently suggested in fault detection [6, 7] and adaptive control [8] problems. RSMI methods are mostly inspired by the offline versions of subspace model identification (SMI) techniques. Now, one of the reasons for the success of SMI techniques lies in the direct correspondence between geometric operations on matrices constructed from input-output data and their implementation in terms of well known, stable and reliable algorithms from the field of numerical linear algebra. However, most of the available batch SMI techniques are based on tools such as the singular value decomposition (SVD) which are not suitable for online implementation due to their computational complexity. In fact, in many online identification scenarios it is important to update the model as time goes on with a reduced computational cost. Thus, the challenging problem with RSMI is to develop new techniques avoiding the use of such burdensome tools. Furthermore, an unbiased updating scheme must be worked out in the presence of both measurement and process noise. Historically, two main approaches to circumvent these difficulties can be distinguished. First, some works proposed adaptations of SMI algorithms in order to update the SVD [1, 2]. Unfortunately, these algorithms had the drawback of demanding that the disturbances acting on the system output were spatially and temporally white, which is obviously restrictive in practice. The second approach [3, 4, 5, 9] relied on the similarities between RSMI and signal processing techniques for direction of arrival (DOA) estimation and exploited the so-called Yang's criterion [10]. In particular, DOA estimation algorithms were adapted in order to deal with more general types of perturbations than the ones arising in the DOA framework, thanks to the use of instrumental variables (IV). Thus, in [3], it was first suggested to resort to an IV adaptation of the projection approximation subspace tracking (PAST) technique [10] to the problem of RSMI, with the focus on computational efficiency. In [9], recursive versions of the so-called PI and PO MOESP SMI schemes and an extension of the above recursive techniques to the case of non linear models of the Wiener type were developed, the main point being the estimation accuracy. [4] is an interesting overview of the previous contributions. More recently, in [5], the recursive estimation of the observability matrix was considered from the initial fourth-order criterion of Yang [10], with a gradient type minimization technique. A convergence proof of the gradient iteration was also provided. Thus, all of the above cited

RSMI algorithms apply certain updating techniques that avoid application of the SVD while preserving the unbiasedness of the original batch algorithms.

The aim of this paper is to present some recent developments in the RSMI class of algorithms [11, 12] for the update of the observability subspace of the model. More precisely, the methods proposed herein are based on the application to system identification of two particular tracking schemes (respectively [13] and [14]), for which an IV version has also been developed in order to take into account the presence of input-output measurement disturbances and process noise in the problem formulation. The main advantage of these approaches in comparison with the PAST subspace tracking technique and its IV variants used in [3] is that no approximation is introduced in the formulation of the tracking problem. In fact, in both cases, the signal subspace is interpreted as the solution of a specific unconstrained optimisation problem.

The paper is organised as follows: Section 2 presents the necessary background about the MOESP class of SMI algorithms before the development of the corresponding recursive versions. Section 3 introduces the analogy between the problem of the estimation of the observability matrix in RSMI and signal subspace tracking in array signal processing. From this relation, the stages necessary to the recursive estimation of the state-space matrices are stressed on and respectively developed in Section 4 and 5. More precisely, in Section 4, the update of the observation vector from the new input-output measurements is proposed thanks to the use of some specific RQ factorization updates. In Section 5, two novel approaches to the estimation of a basis of the observability subspace are proposed and analysed. In Section 6, the performance of these two new techniques is analyzed from numerical simulations. Section 7 concludes the paper.

Throughout the paper, the following notations are used: capital letters are employed for matrices, roman letters for stacked vectors. As usual, the exponent ^{T} and ^{H} respectively denote the transposition and the Hermitian transposition. Each matrix or vector crowned with ˆ is the estimate of the considered matrix or vector. $E[.]$ denotes the mathematical expectation operator, $\|.\|$ the Euclidean norm, $\|.\|_F$ the Frobenius norm and $Im_{col}(.)$ defines the subspace spanned by the columns of the studied matrix. $\delta(s,t)$ denotes the Kronecker delta function.

2 Overview of the MOESP algorithms

Consider an $n-th$ order causal linear time-invariant state-space model with l outputs and m inputs, respectively $\tilde{y}$ and $\tilde{u}$:

$$\begin{array}{rcl} x(t+1) & = & Ax(t) + B\tilde{u}(t) + w(t) \\ \tilde{y}(t) & = & Cx(t) + D\tilde{u}(t) \end{array} \tag{1}$$

where $w \in \mathbb{R}^n$ is the process noise. The measured input and output signals are modeled as

$$\begin{array}{l} u(t) = \tilde{u}(t) + f(t) \\ y(t) = \tilde{y}(t) + v(t) \end{array} \tag{2}$$

with $f \in \mathbb{R}^m$ and $v \in \mathbb{R}^l$ the measurement noises (see Figure 1). Each noise is assumed to be zero-mean white noise and statistically independent of the past noise-free input $\tilde{u}$. Furthermore, the measurement noises f and v are assumed to be independent of the state x.
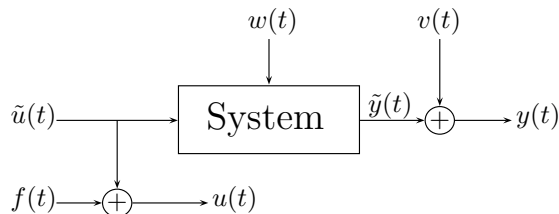

Figure 1: Block schematic illustration of the studied system.

The key problem dealt with by the MOESP class of identification algorithms is the consistent estimation of the column space of the extended observability matrix Γ_i, defined as

$$\Gamma_i = \begin{bmatrix} C^T & (CA)^T & (CA^2)^T & \dots & (CA^{i-1})^T \end{bmatrix}^T \tag{3}$$

2

from measured input-output samples $\{u(t), y(t)\}$. Indeed, from the estimate of the column space of Γ_i, an estimate of the matrices A and C can be derived (up to a similarity transformation) in a straightforward way by exploiting the shift invariance of the column space of Γ_i. From then on, the estimation of B and D can be performed in a number of ways, for example by minimising the simulation error over the identification data set, thus leading to a conventional linear least squares problem (see e.g. [15]) which can also be recursively updated as described in [3].

The starting point for the estimation of the column space of Γ_i is the so called "data equation". To define this equation, introduce the following block-Hankel matrices

$$Y_{t,i,j} = \begin{bmatrix} y(t) & \cdots & y(t+j-1) \\ y(t+1) & \cdots & y(t+j) \\ \vdots & \ddots & \vdots \\ y(t+i-1) & \cdots & y(t+i+j-2) \end{bmatrix} U_{t,i,j} = \begin{bmatrix} u(t) & \cdots & u(t+j-1) \\ u(t+1) & \cdots & u(t+j) \\ \vdots & \ddots & \vdots \\ u(t+i-1) & \cdots & u(t+i+j-2) \end{bmatrix} \tag{4}$$

where i and j are user defined integers chosen such that $j >> i > n$. Then, the data equation is compactly denoted as [16]

$$Y_{t,i,j} = \Gamma_i X_{t,j} + H_i U_{t,i,j} + \underbrace{G_i W_{t,i,j} - H_i F_{t,i,j} + V_{t,i,j}}_{B_{t,i,j}} \tag{5}$$

where $W_{t,i,j}$, $F_{t,i,j}$ and $V_{t,i,j}$ are defined in the same way than $U_{t,i,j}$ and $Y_{t,i,j}$. H_i and G_i are given by

$$H_i = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{i-2}B & CA^{i-3}B & \cdots & D \end{bmatrix} G_i = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ CA & C & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{i-2} & CA^{i-3} & \cdots & 0 \end{bmatrix} \tag{6}$$

and

$$X_{t,j} = \begin{bmatrix} x(t) & x(t+1) & \cdots & x(t+j-1) \end{bmatrix}. \tag{7}$$

On the basis of this data equation, a number of versions of MOESP were developed in order to work out accurate estimates of the column space of the observability matrix from the available input-output data. The algorithms mainly differ according to the assumptions on the disturbances acting on the system generating the data. A summary of the considered MOESP identification algorithms is given in the following. Note that the consistency results corresponding to each of the considered algorithms implicitly assume appropriate persistency of excitation conditions (see the cited papers for details).

The elementary MOESP (EM) scheme [17]: this scheme relies on the (restrictive) assumption that an estimate of the system matrices $[A, B, C, D]$ (up to a similarity transformation) is available. In this case, an estimate $\hat{H}_i$ of H_i can be constructed. Then, the Hankel input-output data matrices verify

$$Z_{t,i,j} = Y_{t,i,j} - \hat{H}_i U_{t,i,j} = \Gamma_i X_{t,j} + B_{t,i,j} + (H_i - \hat{H}_i)U_{t,i,j}. \tag{8}$$

Assume that f, w equal zero and v is a zero-mean white noise independent of the input u. Then, if a consistent estimate of $[A, B, C, D]$ is available, a consistent estimate of $Im_{col}(\Gamma_i)$ can be obtained via an SVD of $Z_{t,i,j}$.

The ordinary MOESP (OM) scheme [17]: in this case, no a priori estimate of the quadruple $[A, B, C, D]$ is necessary. The OM scheme considers the RQ factorization of the compound matrix

$$\begin{bmatrix} U_{t,i,j} \\ Y_{t,i,j} \end{bmatrix} = \begin{bmatrix} R_{11}(\bar{t}) & 0 \\ R_{21}(\bar{t}) & R_{22}(\bar{t}) \end{bmatrix} \begin{bmatrix} Q_1(\bar{t}) \\ Q_2(\bar{t}) \end{bmatrix}, \quad \begin{bmatrix} Q_1(\bar{t}) \\ Q_2(\bar{t}) \end{bmatrix} \begin{bmatrix} Q_1^T(\bar{t}) & Q_2^T(\bar{t}) \end{bmatrix} = I \tag{9}$$

with $\bar{t} = t + i + j - 2$. Then, a consistent estimate of the column space of Γ_i is provided via an SVD of the matrix $R_{22}(\bar{t})$ under the same assumptions on the perturbations stated for the EM scheme.

Remark 1. *Both the EM and OM schemes try to remove the term $H_i U_{t,i,j}$ from $Y_{t,i,j}$. The difference lies in the way the matrix H_i is estimated. In the first case, $\hat{H}_i$ is constructed from consistent estimates of $[A, B, C, D]$. In the second, the estimation of H_i is made from the following projection calculated via the RQ factorization (9) [18]:*

$$Y_{t,i,j}\Pi_U^\perp = Y_{t,i,j} - \hat{H}_i^{LS}U_{t,i,j} = R_{22}(\bar{t})Q_2(\bar{t}) \tag{10}$$

where $\Pi_U^\perp$ denotes the orthogonal projection onto the nullspace of $U_{t,i,j}$ and $\hat{H}_i^{LS}$ the unconstrained minimizing argument of $\|Y_{t,i,j} - H_i U_{t,i,j}\|_F^2$.

The past input past output (PI/PO) MOESP schemes [19]: these schemes consider the following RQ factorization

$$\begin{bmatrix} U_{t+i,i,j} \\ \Xi_{t,i,j} \\ Y_{t+i,i,j} \end{bmatrix} = \begin{bmatrix} R_{11}(\bar{t}) & 0 & 0 \\ R_{21}(\bar{t}) & R_{22}(\bar{t}) & 0 \\ R_{31}(\bar{t}) & R_{32}(\bar{t}) & R_{33}(\bar{t}) \end{bmatrix} \begin{bmatrix} Q_1(\bar{t}) \\ Q_2(\bar{t}) \\ Q_3(\bar{t}) \end{bmatrix} \tag{11}$$

with $\bar{t} = t + 2i + j - 1$ and where $\Xi_{t,i,j} = U_{t,i,j}$ in the PI scheme, while $\Xi_{t,i,j} = [U_{t,i,j}^T\ Y_{t,i,j}^T]^T$ in the PO scheme. In the PI scheme (see [20] for details), a consistent estimate of the column space of Γ_i is provided via an SVD of the matrix $R_{32}(\bar{t})$ under the assumptions that $f \equiv 0$, $w \equiv 0$ and v is an ergodic sequence satisfying $E[u(t)v(s)^T] = 0\ \forall(t,s)$. Likewise, the PO scheme supplies reliable estimates from the same SVD if $f \equiv 0$ and (v, w) are ergodic sequences satisfying

$$E\left[\begin{bmatrix} v(t) \\ w(t) \end{bmatrix} \begin{bmatrix} v(s)^T & w(s)^T \end{bmatrix}\right] = \begin{bmatrix} L & K \\ K^T & M \end{bmatrix} \delta(s,t) \tag{12}$$

which are furthermore independent from the input u. Finally, note that both the PI and the PO schemes can deliver unbiased estimates of the column space of Γ_i in the presence of a zero mean white input measurement vector f provided that the noise free input $\tilde{u}$ is also a realization of zero mean white noise (see [16] for details).

As mentioned in the previous paragraphs, the actual extraction of the column space of Γ_i from respectively $Z_{t,i,j}$, $R_{22}(\bar{t})$ and $R_{32}(\bar{t})$, can be carried out in a very simple way by means of a singular value decomposition of these matrices. While this task is easily accomplished in a batch implementation of the identification algorithms, when the problem of recursive operation is considered, the SVD turns out to be the most burdensome step of MOESP from a computational point of view.

In many online applications, it is essential to perform a recursive adaptation of the model parameters: it is then necessary to develop efficient techniques in order to update the matrix estimates.

3 Recursive subspace identification

The problem we are concerned with is the *recursive* estimation of a state-space realization from the updates of the perturbed input-output data $u(t)$ and $y(t)$. In the light of the discussion of the previous Section, the main issue of the recursive implementation of the MOESP class is the recursive update of Γ_i. Thus, it is essential to define a suitable and efficient computational scheme for the extraction of its column space, without performing the full singular value decomposition at each time step. To this purpose, consider the Hankel matrices (4) and assume that, at time $\bar{t}+1$, new data samples $y(\bar{t}+1)$ and $u(\bar{t}+1)$ are acquired. Then, each of the previous matrices are modified by the addition of a column, which is respectively denoted with $y_i(\bar{t}+1)$ and $u_i(\bar{t}+1)$

$$y_i(\bar{t}+1) \quad = [y^T(t+j)\cdots y^T(t+i+j-1)]^T \in \mathbb{R}^{li\times 1} \tag{13}$$

$$u_i(\bar{t}+1) \quad = [u^T(t+j)\cdots u^T(t+i+j-1)]^T \in \mathbb{R}^{mi\times 1} \tag{14}$$

From the state-space representation (1) and (2), it is easy to show that the output vector is given by

$$y_i(\bar{t}+1) = \Gamma_i x(\bar{t}-i+2) + H_i u_i(\bar{t}+1) + \underbrace{G_i w_i(\bar{t}+1) - H_i f_i(\bar{t}+1) + v_i(\bar{t}+1)}_{b_i(\bar{t}+1)}, \tag{15}$$

4

where the stacked noise vectors are defined in the same way as $u_i(\bar{t}+1)$ and $y_i(\bar{t}+1)$. The main objective is to work out an approach to the estimation of a basis for the column space of matrix Γ_i with a reduced computational cost. In the array signal processing field, several adaptive algorithms were suggested as SVD alternatives to estimate the signal subspace [21, 22]. These techniques are based on the following data generation model

$$z(\bar{t}+1) = \Gamma(\theta)s(\bar{t}+1) + b(\bar{t}+1), \tag{16}$$

where $z(\bar{t}+1) \in \mathbb{C}^{n_z \times 1}$ is the output of the n_z sensors of the antenna array at time $\bar{t}+1$, $\Gamma(\theta) \in \mathbb{C}^{n_z \times n}$ the steering matrix for a direction of arrival vector θ, $s(\bar{t}+1) \in \mathbb{C}^{n \times 1}$ the vector of the n ($n < n_z$) random impinging waves and $b(\bar{t}+1) \in \mathbb{C}^{n_z \times 1}$ the measurement noise. The connection between the RSMI problem and the above model from array signal processing becomes apparent by writing (15) as

$$z_i(\bar{t}+1) = y_i(\bar{t}+1) - H_i u_i(\bar{t}+1) = \Gamma_i x(\bar{t}-i+2) + b_i(\bar{t}+1). \tag{17}$$

The analogy between (16) and (17) is also obvious (see Table 1). This expression stresses on both steps

Subspace Identification	Array Signal Processing
$z_i(\bar{t}+1) \in \mathbb{R}^{li \times 1}$	$z(\bar{t}+1) \in \mathbb{C}^{n_z \times 1}$
$\Gamma_i \in \mathbb{R}^{li \times n}$	$\Gamma(\theta) \in \mathbb{C}^{n_z \times n}$
$x(\bar{t}-i+2) \in \mathbb{R}^{n \times 1}$	$s(\bar{t}+1) \in \mathbb{C}^{n \times 1}$
$b_i(\bar{t}+1) \in \mathbb{R}^{li \times 1}$	$b(\bar{t}+1) \in \mathbb{C}^{n_z \times 1}$

Table 1: Relations between array signal processing and subspace identification

required to recursively estimate the state-space matrices of the system:

- the update of the"observation vector" z_i from the input-output measurements

$$z_i(\bar{t}+1) = y_i(\bar{t}+1) - H_i u_i(\bar{t}+1) \tag{18}$$

- the estimation of a basis of Γ_i from this observation vector

$$z_i(\bar{t}+1) = \Gamma_i x(\bar{t}-i+2) + b_i(\bar{t}+1). \tag{19}$$

These steps are respectively discussed in Section 4 and Section 5.

4 Update of the observation vector

In order to recursively estimate the observability matrix, it is necessary to determine, at each time step, an accurate update of the observation vector $z_i(\bar{t}+1)$. If the matrix $H_i(\bar{t}+1)$ was a priori known, it would be possible to subtract $H_i(\bar{t}+1)u_i(\bar{t}+1)$ from $y_i(\bar{t}+1)$. Unfortunately, this matrix is unknown at time $\bar{t}+1$. Two alternative solutions can be considered:

- to use the state-space matrix estimates available at time $\bar{t}$ to construct $\hat{H}_i(\bar{t})$ [3] and introduce the following approximation

$$\check{z}_i(\bar{t}+1) \approx y_i(\bar{t}+1) - \hat{H}_i(\bar{t})u_i(\bar{t}+1) \tag{20}$$

- to avoid the above approximation and compute a quantity which contains the same information as $z_i(\bar{t}+1)$ by updating the RQ factorisations (9) and (11) [1, 9, 23].

This second method is developed in the following Subsections (see also [4]), the first one, named "subtraction method", having already been used in [3, 4, 12].

4.1 The OM update

The basic idea of this approach is to directly update the RQ factorization of the OM scheme at each time step. For that purpose, assume that the RQ factorization of the input-output Hankel matrices at time $\bar{t} = t + i + j - 2$ is given by equation (9). At each new sample time, this decomposition can be updated as

$$\begin{bmatrix} U_{t,i,j+1} \\ Y_{t,i,j+1} \end{bmatrix} = \begin{bmatrix} R_{11}(\bar{t}) & 0 & \mathrm{u}_i(\bar{t}+1) \\ R_{21}(\bar{t}) & R_{22}(\bar{t}) & \mathrm{y}_i(\bar{t}+1) \end{bmatrix} \begin{bmatrix} Q_1(\bar{t}) & 0 \\ Q_2(\bar{t}) & 0 \\ 0 & 1 \end{bmatrix}. \tag{21}$$

A sequence of Givens rotations $G(\bar{t}+1)$ [24] can then be used to annihilate the vector $\mathrm{u}_i(\bar{t}+1)$ and bring back the R factor to block lower triangular form

$$\begin{bmatrix} R_{11}(\bar{t}) & 0 & \mathrm{u}_i(\bar{t}+1) \\ R_{21}(\bar{t}) & R_{22}(\bar{t}) & \mathrm{y}_i(\bar{t}+1) \end{bmatrix} G(\bar{t}+1) = \begin{bmatrix} R_{11}(\bar{t}+1) & 0 & 0 \\ R_{21}(\bar{t}+1) & R_{22}(\bar{t}) & \bar{\mathrm{z}}_i(\bar{t}+1) \end{bmatrix}. \tag{22}$$

$G(\bar{t}+1)$ is here the matrix ensued from the product of the successive Givens rotations necessary to zero out $\mathrm{u}_i(\bar{t}+1)$. $\bar{\mathrm{z}}_i(\bar{t}+1)$ is the vector obtained by modifying $\mathrm{y}_i(\bar{t}+1)$ in order to include the information contained in $\mathrm{u}_i(\bar{t}+1)$ and in the block $\begin{bmatrix} R_{11}^T(\bar{t}) \ R_{21}^T(\bar{t}) \end{bmatrix}^T$.

Since the goal of this calculation is to evaluate equation (18), it is essential to find the existing relation between $\bar{\mathrm{z}}_i(\bar{t}+1)$ and $\mathrm{z}_i(\bar{t}+1)$. Such relation is defined by the following Proposition.

Proposition 1. *Consider the definition of the observation vector $\mathrm{z}_i(\bar{t}+1)$ (equation (18)) and the vector $\bar{\mathrm{z}}_i(\bar{t}+1)$ computed with the Givens rotations OM update (22). Then, $\bar{\mathrm{z}}_i(\bar{t}+1) = \pm \mathrm{z}_i(\bar{t}+1)$.*

Proof. See Appendix A. $\qquad\square$

On the basis of Proposition 1, the explicit calculation of (18) can be carried out at each time step from an update of the RQ factorization of the block matrix $\begin{bmatrix} Y_{t,i,j}^T \ U_{t,i,j}^T \end{bmatrix}^T$ via an appropriate sequence of Givens rotations.

4.2 The PI/PO update

The main idea in this algorithm is to update the more complex RQ factorization of the PI/PO schemes by means of Givens rotations. In this Subsection, the application of this approach to the PI version of MOESP is proposed (see [4] for details and an illustration of the PO update).

Consider the PI MOESP described in Section 2. At the generic time instant $\bar{t}$, the RQ factorization of the data matrix is given by (11) where, in this case, $\bar{t} = t + j + 2i - 1$. Assume now that the following new set of input-output data vectors becomes available

$$\mathrm{u}_i^f(\bar{t}+1) = \begin{bmatrix} u(t+i+j)^T & \ldots & u(t+2i+j-1)^T \end{bmatrix}^T \tag{23}$$

$$\mathrm{y}_i^f(\bar{t}+1) = \begin{bmatrix} y(t+i+j)^T & \ldots & y(t+2i+j-1)^T \end{bmatrix}^T \tag{24}$$

and define the vector of past input data available at time $\bar{t}+1$ as

$$\mathrm{u}_i^p(\bar{t}+1) = \begin{bmatrix} u(t+j)^T & \ldots & u(t+i+j-1)^T \end{bmatrix}^T. \tag{25}$$

Then, a new column is added to the data matrices and the decomposition must be written as

$$\begin{bmatrix} R_{11}(\bar{t}) & 0 & 0 & \mathrm{u}_i^f(\bar{t}+1) \\ R_{21}(\bar{t}) & R_{22}(\bar{t}) & 0 & \mathrm{u}_i^p(\bar{t}+1) \\ R_{31}(\bar{t}) & R_{32}(t) & R_{33}(\bar{t}) & \mathrm{y}_i^f(\bar{t}+1) \end{bmatrix} \begin{bmatrix} Q_1(\bar{t}) & 0 \\ Q_2(\bar{t}) & 0 \\ Q_3(\bar{t}) & 0 \\ 0 & 1 \end{bmatrix}. \tag{26}$$

Givens rotations are then used twice to update the factorization. They are first applied in order to zero out the elements of vector $\mathrm{u}_i^f(\bar{t}+1)$, bringing the R factor to the form

$$\begin{bmatrix} R_{11}(\bar{t}+1) & 0 & 0 & 0 \\ R_{21}(\bar{t}+1) & R_{22}(\bar{t}) & 0 & \bar{\mathrm{z}}_i^p(\bar{t}+1) \\ R_{31}(\bar{t}+1) & R_{32}(\bar{t}) & R_{33}(\bar{t}) & \bar{\mathrm{z}}_i^f(\bar{t}+1) \end{bmatrix}. \tag{27}$$

6

Subsequently, the elements of $\bar{z}_i^p(\bar{t}+1)$ are zeroed in a similar way, to give

$$\begin{bmatrix} R_{11}(\bar{t}+1) & 0 & 0 & 0 \\ R_{21}(\bar{t}+1) & R_{22}(\bar{t}+1) & 0 & 0 \\ R_{31}(\bar{t}+1) & R_{32}(\bar{t}+1) & R_{33}(\bar{t}) & \bar{\bar{z}}_i^f(\bar{t}+1) \end{bmatrix}. \tag{28}$$

Then it is easy to show that the "square" of block $R_{32}(\bar{t}+1)$ can be written as

$$R_{32}(\bar{t}+1)R_{32}^T(\bar{t}+1) = R_{32}(\bar{t})R_{32}^T(\bar{t}) + [\bar{z}_i^f(\bar{t}+1)][\bar{z}_i^f(\bar{t}+1)]^T - [\bar{\bar{z}}_i^f(\bar{t}+1)][\bar{\bar{z}}_i^f(\bar{t}+1)]^T. \tag{29}$$

Thus, in this case, the subspace estimate at time $\bar{t}+1$ is related to the one at time $\bar{t}$ via the combination of an update and a downdate.

Remark 2. *Note that the PI/PO scheme, which is possibly the most accurate one, suffers from the highest computational cost. Therefore, the OM scheme (see Subsection 4.1) seems to provide an adequate compromise solution between the subtraction approach (see equation (20)) and the PI/PO update (see Subsection 4.2).*

Remark 3. *It is interesting to point out that schemes for the update of the observation vectors similar to the ones discussed above can be derived by computing explicitly the projections associated with the RQ factorisations computed in the OM and PI/PO cases and applying the matrix inversion lemma to such expressions. To our best knowledge, this approach has been followed independently in [5] for the OM and PI/PO schemes and in [25] for the OM update.*

The first step of the problem, i.e., the reliable update of the observation vector, is now achieved. Henceforth, consider the second stage of the RSMI problem, that is the recursive estimation of the observability matrix.

5 Update of the observability subspace

The most important step in RSMI is the recursive update of the observability subspace. In [3, 4], it was proposed to exploit the close relationship between array signal processing and SMI to derive efficient SVD alternatives. More precisely, the PAST algorithm [10] and its instrumental variables modification IVPAST [3] were applied and modified to derive an effective update of the signal subspace. In particular, an optimisation problem was considered with respect to a fourth-order cost function so as to retrieve the signal subspace from I/O data, the efficient recursive solution of which required the introduction of an approximation, in order to reduce the optimisation task to an exponentially weighted least-squares problem. The aim of this Section is therefore to review in greater detail the PAST and IVPAST techniques and to introduce two novel approaches to the problem of updating the observability subspace of the identified models.

5.1 Overview of PAST and IVPAST

5.1.1 PAST

Consider a random vector $z \in \mathbb{R}^{n_z}$, and study the following *unconstrained* criterion (known as Yang's criterion)

$$V(W) = E \left\| z - WW^T z \right\|^2 \tag{30}$$

with a matrix argument $W \in \mathbb{R}^{n_z \times n}$, $n_z > n$ that, without loss of generality, is assumed to have full rank ($=n$). Let the eigenvalue decomposition of $R_z = E[zz^T]$ be given as

$$R_z = Q\Lambda Q^H \tag{31}$$

with $Q = [q_1, \ldots, q_{n_z}]$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{n_z})$. Furthermore, assume that R_z is positive definite. The eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n_z}$. Then, the following theorem holds [10]:

Theorem 1. *The global minimum of $V(W)$ is attained if and only if $W = Q_n T$ where Q_n contains the n dominating eigenvectors of R_z and T is an arbitrary unitary matrix. Furthermore, all other stationary points are saddle points.*

This theorem allows to formulate the problem of computing the n dominating eigenvectors of a positive definite matrix as an unconstrained optimization problem. In particular, if the expectation operator in (30) is replaced with a finite summation and a suitable forgetting factor λ is introduced, the cost function V becomes

$$V(W(t)) = \sum_{k=1}^{t} \lambda^{t-k} \left\| \mathrm{z}(k) - W(t)W^T(t)\mathrm{z}(k) \right\|^2 \tag{32}$$

and it is possible to derive a practical recursive algorithm, along the lines described below. Theorem 1 is also applicable to (32) if R_z is replaced with

$$\hat{R}_z(t) = \sum_{k=1}^{t} \lambda^{t-k} \mathrm{z}(k)\mathrm{z}^T(k). \tag{33}$$

The key idea of PAST is to replace $W^T(t)\mathrm{z}(k)$ in (32) with

$$\mathrm{h}(k) = W^T(k-1)\mathrm{z}(k). \tag{34}$$

This so-called *projection approximation* results in the criterion

$$\bar{V}(W(t)) = \sum_{k=1}^{t} \lambda^{t-k} \left\| \mathrm{z}(k) - W(t)\mathrm{h}(k) \right\|^2 \tag{35}$$

which is quadratic in $W(t)$, and is minimized by

$$W(t) = \hat{R}_{\mathrm{zh}}(t)\hat{R}_{\mathrm{h}}^{-1}(t) \tag{36}$$

with obvious definitions of the involved matrices. It is further assumed that the involved inverse exists. When the matrix inversion lemma is applied to (36), an efficient RLS-like algorithm is easily derived (see [10]).

Remark 4. *With the PAST criterion, the signal subspace is obtained by minimizing the modified cost function $\bar{V}(W)$ instead of $V(W)$. Hence, the estimated column-subspace is slightly different from the one reachable with the original cost function. Theoretically, the columns of W minimizing the criterion $V(W)$ are orthonormal. Even if this property is not necessary to extract the state-space matrices, the minimization of $\bar{V}(W)$ leads to a matrix having columns that are not exactly orthonormal. This property evolves during the recursive minimization since, under some conditions, the minimizer of $\bar{V}(W)$ converges to a matrix with orthonormal columns [10]. This evolution can be interpreted as a slow change of basis, which implies that it is not possible to guarantee that $\Gamma_i(t)$ and $\Gamma_i(t-1)$ are expressed in the same state-space coordinates. This might represent a problem in the estimation of the state space realization $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$, and particularly so whenever RSMI techniques are used for change detection purposes (see, e.g., [6, 7]).*

Remark 5. *In [10] a gradient based approach to the optimisation of (30) was also proposed. Indeed, the gradient of $V(W)$ with respect to W has the simple expression*

$$\nabla V(W) = [-2R_z + R_z WW^T + WW^T R_z]W, \tag{37}$$

which can be used to implement a gradient update rule (either iterative or recursive) for the subspace estimate. This approach has been followed in [5] for the update of the observability subspace.

5.1.2 IVPAST

An IV generalization of PAST has been proposed in [3], in order to deal with situations in which the measurements of z are affected by noise with arbitrary and unknown covariance matrix. In this scenario, it is assumed that the cross correlation matrix $R_{z\xi}$ associated with z and an instrumental variable vector $\xi \in \mathbb{R}^{\gamma \times 1}$ has a low rank $(= n)$ structure

$$E[z\xi^T] = \Gamma\Phi \tag{38}$$

where $\Gamma \in \mathbb{R}^{n_z \times n}, \Phi \in \mathbb{R}^{n \times \gamma}$ both have full rank n $(n_z, \gamma \geq n)$. Then, considering the criterion (due to Gustafsson)

$$V_{IV}(W(t)) = \left\| R_{z\xi}(t) - W(t)W^T(t)R_{z\xi}(t) \right\|_F^2 \tag{39}$$

the following result can be introduced [3]:

Theorem 2. *Let $R_{z\xi}(t)$ have the SVD*

$$R_{z\xi}(t) = Q\Sigma V^T. \tag{40}$$

Then the global minimum of $V_{IV}(W(t))$ is obtained if and only if $W(t) = Q_n T$ where Q_n contains the n dominating left singular vectors of $R_{z\xi}(t)$ and T is an arbitrary unitary matrix. All other stationary points are saddle points.

Applying the projection approximation approach, a recursive algorithm can be derived (see [3]).

As was pointed out in Remark 4, the PAST algorithm and its IV versions are based on an approximation. In order to avoid this difficulty, two novel algorithms, inspired from array signal processing, are respectively introduced in Subsections 5.2 and 5.3.

5.2 RSMI using the Propagator Method

5.2.1 The Propagator Method

The propagator [14] is a linear operator used in array signal processing to provide the decomposition of the observation space into a noise and a signal subspace. In order to understand the key point of this method, consider again the sensor array signal model (16). Assuming that the steering matrix Γ is of full rank, n rows of Γ are linearly independent, the others being expressed as a linear combination of these n rows. Under this hypothesis, and after a reorganization of the sensors outputs so that the first n rows of Γ are linearly independent, it is possible to partition the steering matrix according to

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \begin{array}{l} \} \in \mathbb{C}^{n \times n} \\ \} \in \mathbb{C}^{n_z - n \times n} \end{array} . \tag{41}$$

The propagator [14] is the unique linear operator $P \in \mathbb{C}^{n \times n_z - n}$ defined as

$$\Gamma_2 = P^H \Gamma_1. \tag{42}$$

From [12] and [26], on the basis of the similarity between (16) and (17) (see Table 1) and under the hypotheses that the order n is a priori known and that $\{A, C\}$ is observable, the extended observability matrix can be decomposed in the following way

$$\Gamma_i = \begin{bmatrix} \Gamma_{i_1} \\ \Gamma_{i_2} \end{bmatrix} \begin{array}{l} \} \in \mathbb{R}^{n \times n} \\ \} \in \mathbb{R}^{li - n \times n} \end{array} \tag{43}$$

where Γ_{i_1} is the block of the n independent rows and Γ_{i_2} the matrix of the $li - n$ others. Thus, there is a unique $P \in \mathbb{R}^{n \times li - n}$ such as

$$\Gamma_{i_2} = P^T \Gamma_{i_1}. \tag{44}$$

It is also easy to verify that

$$\Gamma_i = \begin{bmatrix} \Gamma_{i_1} \\ \Gamma_{i_2} \end{bmatrix} = \begin{bmatrix} I_n \\ P^T \end{bmatrix} \Gamma_{i_1} = Q_s \Gamma_{i_1}. \tag{45}$$

This means that the columns of Γ_i are linear combinations of the columns of Q_s. Thus

$$Im_{col}\{\Gamma_i\} = Im_{col}\{Q_s\} \tag{46}$$

since $\text{rank}(\Gamma_{i_1}) = n$. Equation (46) implies, from the knowledge of the propagator P, the ability to find an expression of the observability matrix in a particular basis. Thus, assuming that the order n is known, an estimate of the subspace spanned by the observability matrix is available by estimating P. For that purpose, consider equation (17). After an initial reorganization such that the first n rows of Γ_i are linearly independent, the following partition of the observation vector (18) can be introduced

$$z_i(t) = \begin{bmatrix} I_n \\ P^T \end{bmatrix} \Gamma_{i_1} x(t) + b_i(t) = \begin{bmatrix} z_{i_1}(t) \\ z_{i_2}(t) \end{bmatrix} \begin{array}{l} \} \in \mathbb{R}^{n \times 1} \\ \} \in \mathbb{R}^{li-n \times 1} \end{array} \tag{47}$$

where z_{i_1} and z_{i_2} are the components of z_i corresponding respectively to the n rows of Γ_{i_1} and $li - n$ rows of Γ_{i_2} (the same symbols are used before and after the reorganization, for the sake of simplicity). In the ideal, noise-free case, it is easy to show that

$$z_{i_2} = P^T z_{i_1}. \tag{48}$$

In the presence of noise, this relation holds no longer. An estimate of P can however be obtained by minimizing the following cost function

$$\tilde{V}(P) = E\|z_{i_2} - P^T z_{i_1}\|^2, \tag{49}$$

the uniqueness of $\hat{P}$ being ensured by the convexity of this criterion, which, in turn, can be guaranteed by suitable persistency of excitation assumptions.

The criterion (49) is, by definition, quadratic and reduces the determination of the range of Γ_i to the estimation of an $li - n \times n$ matrix. For example, on the basis of the signal model given in equation (47), under the assumption that

$$E\left[b_i b_i^T\right] = R_{b_i} = \begin{bmatrix} R_{b_{i_1}} & R_{b_{i_1} b_{i_2}} \\ R_{b_{i_2} b_{i_1}} & R_{b_{i_2}} \end{bmatrix}, \tag{50}$$

the (asymptotic) LS estimate of P^T is given by ([27])

$$\hat{P}^T = R_{z_{i_2} z_{i_1}} R_{z_{i_1}}^{-1} \tag{51}$$

where

$$R_{z_{i_2} z_{i_1}} = E\left[z_{i_2} z_{i_1}^T\right] = \left[P^T \Gamma_{i_1} R_x \Gamma_{i_1}^T + R_{b_{i_2} b_{i_1}}\right] \tag{52}$$

and

$$R_{z_{i_1}} = E\left[z_{i_1} z_{i_1}^T\right] = \left[\Gamma_{i_1} R_x \Gamma_{i_1}^T + R_{b_{i_1}}\right]. \tag{53}$$

Therefore, it is easy to see that the LS solution of the optimization problem defined by (49) leads to a biased estimate, even when the residual vector $b_i(t)$ is spatially and temporally white, i.e., when $R_{b_i} = \sigma_b^2 I_{li}$ and therefore

$$\hat{P}^T = P^T \left[\Gamma_{i_1} R_x \Gamma_{i_1}^T\right] \left[\Gamma_{i_1} R_x \Gamma_{i_1}^T + \sigma_b^2 I_n\right]^{-1}. \tag{54}$$

This difficulty could be partially circumvented by considering a Total Least Squares (TLS, see [28]) approach to the problem. However this would likely complicate the recursive implementation of the solution. In order to obtain an unbiased estimate of the signal subspace, an instrumental variable is introduced in the above criterion so as to make it applicable even if colored disturbances act on the system, as done in [3].

5.2.2 Instrumental variable identification using the Propagator Method

Since the propagator method is very sensitive to the presence of the noise b_i, it is necessary to modify the criterion (49). This correction is realized by introducing an instrumental variable $\xi(t) \in \mathbb{R}^{\gamma \times 1}$ ($\gamma \geq n$) in (49), assumed to be uncorrelated with the noise but sufficiently correlated with the state vector $x(t)$, and by defining the new cost function

$$\tilde{V}_{IV}(\hat{P}) = E\|z_{i_2}\xi^T - \hat{P}^T z_{i_1}\xi^T\|_F^2. \tag{55}$$

Since only a finite number of data is accessible in practice, replacing the expectation operator with a finite exponentially weighted sum, criterion (55) becomes

$$\tilde{V}_{IV}(P(t)) = \sum_{k=1}^{t} \lambda^{t-k} \|z_{i_2}(k)\xi^T(k) - P^T(t)z_{i_1}(k)\xi^T(k)\|_F^2. \tag{56}$$

Assuming a suitable instrumental variable can be constructed [12], the minimization of (56) can be realized in two different ways (named IVPM and EIVPM), according to the number γ of instruments in ξ. Indeed, if $\gamma = n$, it is possible to find a first IVPM algorithm by applying the matrix inversion lemma to the least squares solution of the optimisation problem associated with criterion (55)

$$K(t) = \frac{\xi^T(t)R(t-1)}{\lambda + \xi^T(t)R(t-1)z_{i_1}(t)} \tag{57a}$$

$$P^T(t) = P^T(t-1) + [z_{i_2}(t) - P^T(t-1)z_{i_1}(t)]K(t) \tag{57b}$$

$$R(t) = \frac{1}{\lambda}[R(t-1) - R(t-1)z_{i_1}(t)K(t)] \tag{57c}$$

where $R(t) = \{E[z_{i_1}(t)\xi^T(t)]\}^{-1} = R_{z_{i_1}\xi}^{-1}(t)$.

In [29], it was argued that the accuracy of the estimates obtained from an instrumental variable method increases with the number of instruments. It would be interesting to improve the efficiency of the previous algorithm by increasing the number of used instruments (i.e., $\gamma > n$). In that case, the minimization of the criterion (56) appeals to a technique named the Extended Instrumental Variable Method [30]. The application of such a technique gives the recursive updating formulae available in [12] (see also Appendix C), named EIVPM, the main step being

$$P^T(t) = P^T(t-1) + (g(t) - P^T(t-1)\Psi(t))K(t). \tag{58}$$

Remark 6. *From the second iteration on, the update of the estimated subspace is always in the form*

$$\hat{\Gamma}_i = \begin{bmatrix} I_n \\ \hat{P}^T \end{bmatrix}. \tag{59}$$

This means that, after a short transient period, the recursive estimation is made in the same state-space basis. This property is an important asset for the extraction of the state-space matrices.

5.3 RSMI based on Projector Tracking

5.3.1 Projector Tracking based subspace identification

As illustrated in [13], a different approach to the solution of the optimization problem associated with the cost function (30) can be obtained by considering as independent variable the projection matrix $\Pi = WW^T$ rather than the orthonormal basis W. In particular, it is easy to see that optimising $V(W)$ in (30) with respect to W is equivalent to optimising

$$\tilde{\tilde{V}}(\Pi) = E\|z - \Pi z\|^2 = tr(R_z) - tr(R_z\Pi) \tag{60}$$

with respect to Π, with the constraints $\Pi = \Pi^T = \Pi^2$ and subject to the rank constraint rank$(\Pi) = n$. Tracking of a real, $n_z \times n_z$ projection matrix Π of rank n can be carried out by means of the update rule

$$\Pi(\phi, t) = G(\phi, t)\Pi(t-1)G(\phi, t)^T \tag{61}$$

where

$$G(\phi, t) = \prod_{k=n_z-1}^{1} \prod_{l=k+1}^{n_z} G^{kl}(\phi_{kl}) \tag{62}$$

$G^{kl}(\phi_{kl})$ being the Givens rotor (see [24, 31]) with characteristic submatrix

$$G^{kl}\begin{bmatrix} kk & kl \\ lk & ll \end{bmatrix} = \begin{bmatrix} \cos(\phi_{kl}) & -\sin(\phi_{kl}) \\ \sin(\phi_{kl}) & \cos(\phi_{kl}) \end{bmatrix} \tag{63}$$

and $\phi = \{\phi_{kl}\}$. The theoretical foundation for the update rule defined by equation (61) can be found in Theorem 2 and Corollary 1 of [13], which provide the definition for a simple tracking rule for a time-varying projection matrix, in terms of incremental rotations implemented by Givens operators (see also [22]).

At each time step, the elements of vector ϕ, i.e., the angles which represent the projector update, are determined according to the gradient rule

$$\phi_{kl} = -\mu \frac{\partial \tilde{\tilde{V}}(\Pi(\phi))}{\partial \phi_{kl}} \tag{64}$$

where μ is the gain of the gradient iteration. The gradient of $\tilde{\tilde{V}}(\Pi(\phi))$ is given by [13]

$$\frac{\partial \tilde{\tilde{V}}(\Pi(\phi))}{\partial \phi_{kl}} = \sum_{i=1}^{n_z} (\Pi_{l,i} \hat{R}_{\mathsf{z}\,i,k} - \Pi_{k,i} \hat{R}_{\mathsf{z}\,i,l}) \tag{65}$$

where $\Pi_{i,j}$ and $\hat{R}_{\mathsf{z}\,i,j}$ are respectively the (i,j) elements of Π and $\hat{R}_{\mathsf{z}}$.

5.4 Instrumental variable Projector Tracking

In order to provide a more general formulation of the projector tracking approach to RSMI, an instrumental variable version of the algorithm has to be developed. For, note that the cost function (39) can be written in terms of the projection matrix $\Pi = WW^T$ as

$$\tilde{\tilde{V}}_{IV}(\Pi) = \|R_{\mathsf{z}\xi} - \Pi R_{\mathsf{z}\xi}\|_F^2 = tr\left[(R_{\mathsf{z}\xi} - \Pi R_{\mathsf{z}\xi})(R_{\mathsf{z}\xi} - \Pi R_{\mathsf{z}\xi})^T\right] \tag{66}$$

so that letting $R = R_{\mathsf{z}\xi} R_{\mathsf{z}\xi}^T$

$$\tilde{\tilde{V}}_{IV}(\Pi) = tr(R) - tr(R\Pi) \tag{67}$$

which has the same structure as the cost function of the PAST problem (see equation (39)). Therefore, a projector tracking approach can be followed also in the IV case, along the lines of the previous Subsection, thus leading to the definition of instrumental variable subspace identification algorithms based on projection tracking ideas, which in the following will be termed IVPT and EIVPT. In addition, in the light of Theorem 2 in [32], one can work directly with $R_{\mathsf{z}\xi}$ without forming R explicitly. More precisely, taking into account that $\Pi = WW^T$, the recursion for the update of the signal subspace at time t can be outlined as follows:

1. Use the new input-output data to update the RQ factorisation as in Sections 4.1-4.2 or the subtraction as in (20) (possibly using a forgetting mechanism).

2. From a suitable choice of IVs, compute $\hat{R}_{\mathsf{z}\xi} = \mathsf{z}\xi^T$.

3. On the basis of $\hat{R}_{\mathsf{z}\xi} = \mathsf{z}\xi^T$ and the previous estimate $\hat{\Pi}(t-1)$ compute ϕ_{kl} according to

$$\phi_{kl} = -\mu \sum_{i=1}^{n_z} (\hat{\Pi}_{l,i} \hat{R}_{\mathsf{z}\xi\,i,k} - \hat{\Pi}_{k,i} \hat{R}_{\mathsf{z}\xi\,i,l}). \tag{68}$$

4. Update the subspace estimate $\hat{W}(t)$ according to

$$\hat{W}(t) = G(\phi,t)\hat{W}(t-1), \tag{69}$$

with $G(\phi,t)$ as defined in equation (62).

5. Compute matrices A and C, and (possibly) initialise the recursion for B and D.

6. Update the projector estimate $\hat{\Pi}$.

Remark 7. *As pointed out in [13] in the framework of direction of arrival problems for the approach to projector tracking, the algorithm lends itself naturally to a modification for blockwise processing of data, by suitably choosing the number of data that are used at each time step in the estimation of R_z. The same holds for the instrumental variable version of projector tracking proposed herein, in which a blockwise estimate of $R_{z\xi}$ can be used.*

Remark 8. *It has been mentioned in Remark 5 that Yang's criterion (30) (as well as its IV counterpart (39)) can be optimised using a gradient iteration. This approach was applied to RSMI in [5], where a recursive algorithm consisting of a PO update of the observation vector plus a gradient iteration over (30) was proposed. It is interesting to point out the major differences between the projection tracking method proposed herein and the algorithm of [5]. First of all, in [5] the effect of noise is treated at the observation vector level, while in the EIVPT algorithm this is done in the subspace update step. In addition, while the PT update of the observability subspace relies on unitary operations and therefore guarantees the orthonormality of the computed subspace, this is not the case when the estimate is updated using the direct gradient iteration described in Remark 5.*

5.5 Complexity of the propagator and the projector methods

The main goal of recursive identification techniques is to update the estimated model during the operation of the system, at the same time as the data are collected. This condition implies that the identification has to be carried out with a reduced computational cost. Thus, it is interesting to compare the complexity of EIVPM and EIVPT with other existing techniques. Table 2 introduces the overall computational complexity of PAST, EIVPAST, PM, EIVPM, PT and EIVPT. These values concern only the recursive estimation of the observability matrix. It is shown that the computational cost of PM and

Algorithm	Complexity
PAST	$\mathcal{O}(nli)$
EIVPAST	$\mathcal{O}(mli^2)$
PM	$\mathcal{O}(nli)$
EIVPM	$\mathcal{O}(mli^2)$
PT	$\mathcal{O}(mli^2)$
EIVPT	$\mathcal{O}(mli^3)$

Table 2: Overall complexity comparison of several recursive subspace identification algorithms.

EIVPM is comparable to the one of PAST and its by-products, while the current implementation of EIVPT is characterised by a higher computational cost. This complexity, however, does not restrict the online applicability of EIVPT.

5.6 On the equivalence of the propagator and the projector methods

The aim of this Section is to assess the relative merits of the proposed estimation techniques. In particular, two aspects of their performance will be investigated. First of all, since it has been noted in the previous Section that the propagator method without instrumental variables leads to biased estimates, while the projector method, based on the direct optimization of Yang's criterion, is consistent [10, 32], it is necessary to somehow evaluate the importance of the bias introduced by the propagator technique. Subsequently, when coming to the analysis of the instrumental variable versions of both approaches (which are both consistent), it is interesting to check whether the propagator and the projector techniques are equivalent only to first order or also to second order.

5.6.1 Propagator and projector methods

The main result on the relationship between the propagator and the projector methods is given in the following Proposition.

Proposition 2. *The criterion (60) can be equivalently expressed as a function of the propagator P as*

$$\tilde{\tilde{V}}(\Pi) = E\left\|\left[I_n + PP^T\right]^{-1} P\left[P^T z_1 - z_2\right]\right\|^2 + E\left\|\left[P^T P\left[P^T P + I_{li-n}\right]^{-1} - I_{li-n}\right]\left[P^T z_1 - z_2\right]\right\|^2 \quad (70)$$

or

$$\tilde{\tilde{V}}(P) = tr(R_z) - tr(R_z \begin{bmatrix} I_n \\ P^T \end{bmatrix} \left[I_n + PP^T\right]^{-1} [I_n \ P]). \quad (71)$$

Proof. See Appendix B. □

Note that equation (71) can be evaluated numerically as a function of the true and estimated propagator P in order to assess the effect of bias. Precisely, since the LS solution of the propagator estimation problem can be written in closed form (see equations (51)-(53)), the difference between the optimal value of the cost function, which is given by $tr(R_{b_2})$, and the one attained using the propagator method can be computed.

Example 1. *Consider the simple signal model given by*

$$z_i(t) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} x(t) + b_i(t) \quad (72)$$

where x and b_i are (independent) realizations of white noise with zero mean and variance 1 and σ^2 respectively. According to equation (54) the estimate of the propagator is given by

$$\hat{P}^T = \frac{2}{1 + \sigma^2}. \quad (73)$$

In order to assess the performance of the propagator method the following approach has been followed: first, the increase in Yang's cost function with respect to the global minimum has been assessed, along the lines outlined above; then, a Monte Carlo simulation has been carried out, in order to compare the distribution of the estimate of P obtained by the two approaches. In Table 3 the loss of performance associated with the use of the propagator method in a noisy framework is measured in terms of Yang's cost function, for increasing values of the noise variance σ^2. As can be seen from the Table, the normalised performance loss is relatively small[1], in the sense that even in the presence of bias, the estimate of the observability subspace provided by the propagator method still matches fairly well the actual subspace. On the other

Noise variance σ^2	$\frac{\hat{V}-V^o}{V^o}$
0.01	0.008
0.05	0.0392
0.1	0.0768

Table 3: Optimal PAST cost vs optimal cost attained via the propagator method.

hand, the results of the Monte Carlo iterations (a number N=1000 of replications has been carried out) are illustrated in the histograms given in Figure 2. It is interesting to note that while the bias present in the propagator estimates appears clearly, and is obviously increasing with σ^2, the propagator and the PAST estimates of P have very similar second order properties.

5.6.2 IV propagator and projector methods

When considering the instrumental variable versions of the propagator and projector methods, it can be easily concluded, by means of arguments not unlike those given in [29] for the analysis of conventional IV algorithms, that the IVPM and EIVPM algorithms can lead to unbiased estimates of P (and therefore of the observability subspace), i.e., that the introduction of instrumental variables in the propagator approach makes it equivalent (to first order) to all the methods based on Gustafsson's criterion. Unfortunately the comparison between the second order properties of the estimators based on the IVPM and

---

[1]Note, in passing, that one might also use σ^2 as an upper bound for $\frac{\hat{V}-V^o}{V^o}$.
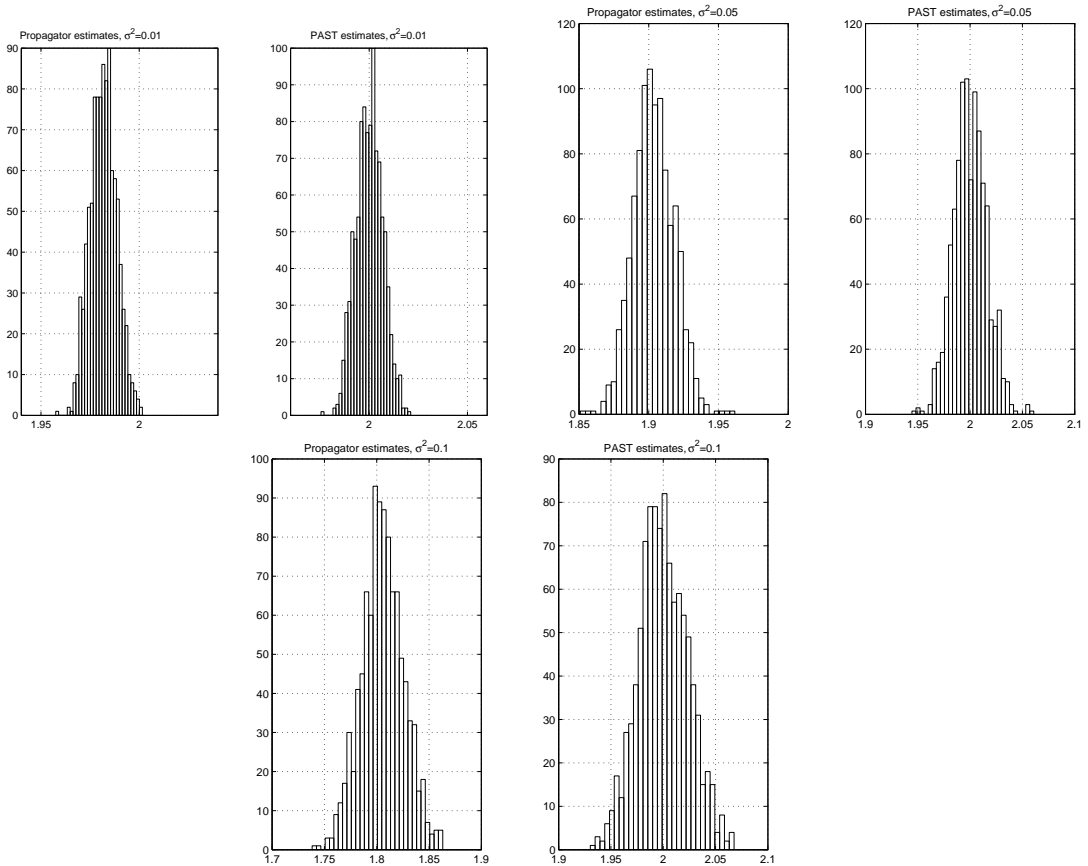
Figure 2: Distribution of the asymptotic estimates of P in Example 1, using the PM and PAST algorithms.

EIVPM approaches is not as simple, and for the time being the only evidence of second order equivalence of the EIVPM and EIVPT methods can be derived experimentally by means of Monte Carlo simulations. Future work will aim at validating theoretically the simulation results illustrated in the following Example 2.

Example 2. *Consider again the data generation model described in Example 1, where each component of the residual* b *is now given by the autoregressive model*

$$\text{b}(t+1) = 0.5\text{b}(t) + \eta(t) \tag{74}$$

where $\eta(t)$ is a realization of zero-mean white Gaussian noise with variance $\sigma^2 = 0.1$. The propagator P has been estimated using both the IVPM and the IVPAST techniques, by choosing as instrumental variable a filtered version of the state sequence x. As in the previous example, $N = 1000$ replications of the estimation procedure for the propagator have been carried out, on the basis of data sets of 200 samples each. The results are summarised in Figure 3, from which it is apparent that the two approaches are entirely equivalent.

Remark 9. *It should be mentioned that the presented recursive schemes can be extended to the PO_EIV errors in variables algorithm and to the estimation of the linear part of Wiener type nonlinear models, along the lines of [4].*

6 Simulation results

In this Section, the performance of the propagator and projector methods are illustrated with two simulation studies. In particular, the identification of a time invariant system and of a time-varying
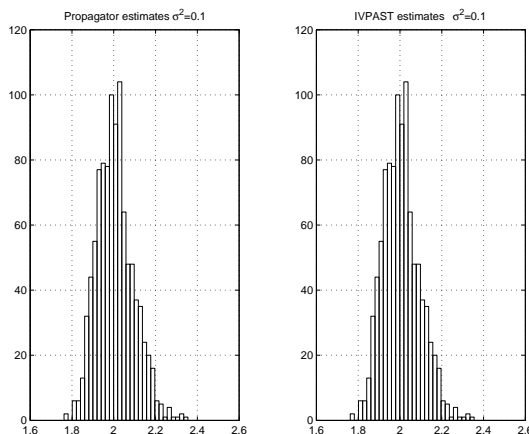
15

Figure 3: Distribution of the asymptotic estimates of P in Example 2, using the IVPM and IVPAST algorithms.

system are considered and the relative merits of the two algorithms are compared and discussed.

6.1 An LTI system

Consider the linear time invariant system described by the equations

$$
\begin{aligned}
x(t+1) &= \begin{bmatrix} 0.7 & 0 \\ 0 & -0.5 \end{bmatrix} x(t) + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tilde{u}(t) + w(t) \\
\tilde{y}(t) &= \begin{bmatrix} 1 & 2 \end{bmatrix} x(t) + 0.05\tilde{u}(t) \\
u(t) &= \tilde{u}(t) + f(t) \\
y(t) &= \tilde{y}(t) + v(t).
\end{aligned}
\tag{75}
$$

The input $\tilde{u}$, a zero-mean white Gaussian noise sequence with variance 1, is contaminated with a zero-mean white Gaussian noise sequence f with variance 0.1. Colored measurement noise v is added to the noise free output $\tilde{y}$, generated as

$$
v(t) = \frac{1}{1 + 0.8q^{-1}} e(t)
\tag{76}
$$

where e is a zero-mean white Gaussian noise with variance 0.1. Finally, a zero-mean white Gaussian process noise w is added with variance 0.1.

The above system has been used in order to carry out a comparison between the algorithms proposed in this paper (EIVPM and EIVPT) and the EIVPAST algorithm (see [3]). For the sake of completeness, two different situations have been considered. In the first one the observation vector z_i is estimated with the subtraction method (see equation (20)), while in the second case, z_i is determined using the OM update (according to Proposition 1). In each situation, the estimated eigenvalues and principal angles between the exact and estimated observability subspace are plotted.

In EIVPM, the forgetting factor is fixed as 0.999. For the projector method, this value is chosen as 0.99, while for EIVPAST a forgetting factor of 0.999 is used[2]. Concerning the initialization of the identification algorithms, the model matrices are randomly generated in a neighbourhood of the true system matrices, i.e., with a possible deviation of 0.3 from the true values of the elements of the matrices. Finally, the i, γ and μ parameters respectively equal 8, 8 and 0.002.

In Figure 4, the results obtained by combining the three observability subspace estimation schemes with the subtraction approach for the update of the observation vector are shown. Similarly, the combination of the three algorithms with the OM update leads to the results presented in Figure 5.

As can be seen from the Figures, all the considered approaches to the RSMI problem yield consistent estimates: the estimated eigenvalues converge to the true ones and the same holds for the estimate of the

---

[2]The values of the forgetting factors have been chosen in order to optimise the performance of each algorithm.
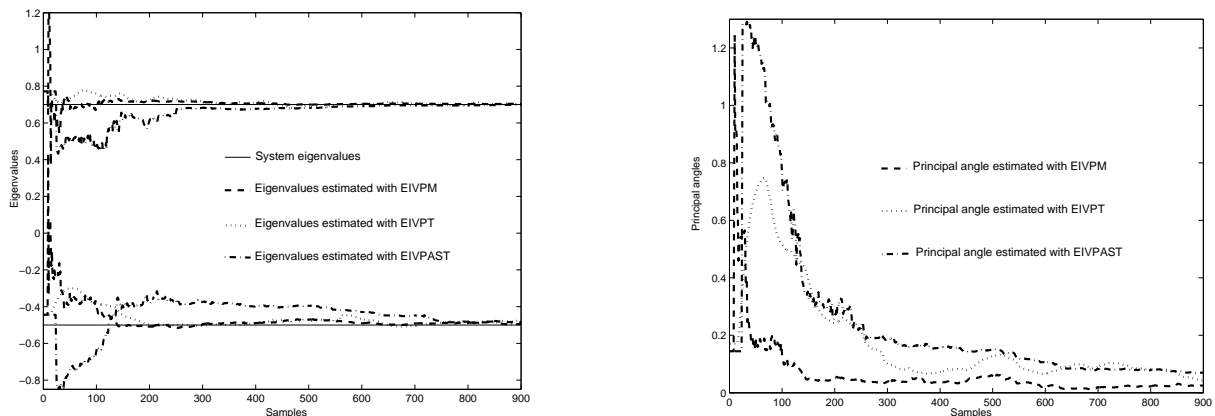
16

Figure 4: Estimated eigenvalues and principal angles calculated with EIVPT, EIVPM and EIVPAST methods using subtraction technique for the estimation of the observation vector. Time invariant case.
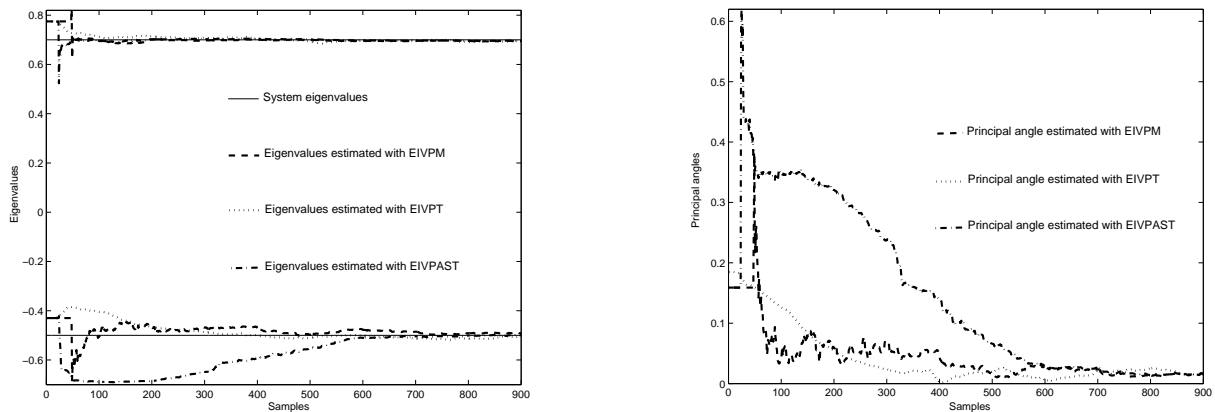

Figure 5: Estimated eigenvalues and principal angles calculated with EIVPT, EIVPM and EIVPAST methods using the OM update technique for the estimation of the observation vector. Time invariant case.

observability subspace. On the other hand, in the considered example, the proposed algorithms show a better performance in terms of transient behaviour with respect to EIVPAST: the slower convergence of EIVPAST is most likely due to the projection approximation.

6.2 A time-varying system

The second simulation study analyses the tracking capabilities of the projector and the propagator methods. For, the above system (75) has been modified in order to present a continuous evolution of its eigenvalues, starting from one third of the simulation duration. In fact, the eigenvalues of A change from $\{-0.5, 0.7\}$ to $\{-0.65, 0.8\}$. In this case, the forgetting factor has been reduced to 0.995 for the EIVPM algorithm, and it has been kept equal to 0.99 in the case of EIVPT. The results (see Figure 6) show that both techniques present interesting abilities to track variations in the system parameters.

6.3 Comments and discussion

During the simulation study it has been noticed that EIVPM and the projector technique each present specific limitations. As far as EIVPM is concerned, under some particular input conditions, the estimation of eigenvalues in the range $[-0.2, 0.2]$ is less accurate. The first studies seem to show that this lack of accuracy is due to the use of the extended instrumental variable technique, this problem being brought
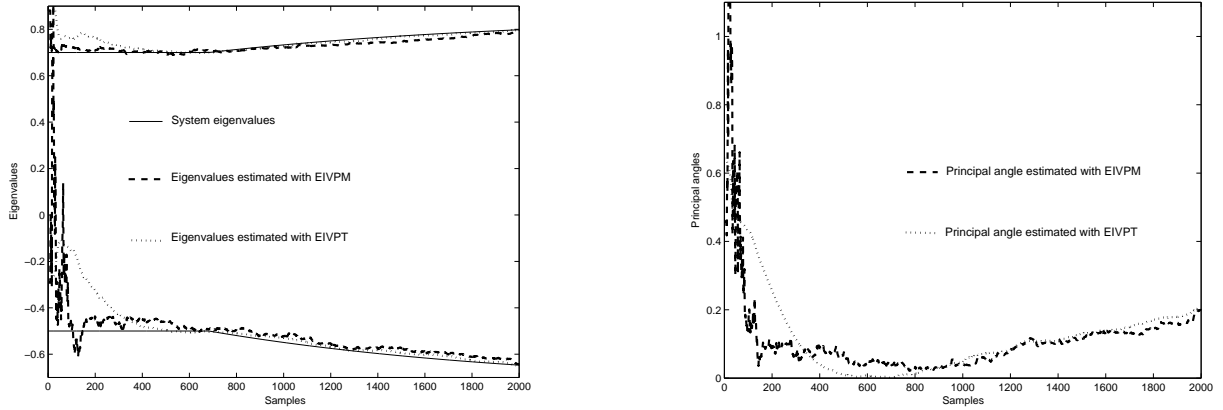
17

Figure 6: Estimated eigenvalues and principal angles calculated with EIVPT, EIVPM and EIVPAST methods using the OM update technique for the estimation of the observation vector. Time varying case.

out with EIVPAST. Some numerical improvements, based on the works of Porat and Friedlander [33], are currently studied. On the other hand, the projector method seems to be dependent on the initialization accuracy. Indeed, convergence problems may arise if the initialization of the system matrices is too far from the true values. This difficulty can be circumvented with a good offline initialization.

7 Concluding remarks

In this paper, two new recursive subspace model identification algorithms, named EIVPM and EIVPT, have been proposed. In both cases, the problem of estimating the observability subspace of the system from input-output data is formulated as a specific unconstrained optimization problem adapted from signal processing. Furthermore, the performance of the projector and propagator methods have been compared with those of the EIVPAST recursive identification algorithm on simulation examples. The good results, both in terms of accuracy and speed of convergence, have been highlighted. Future work will aim at further exploring the issues related with the convergence of the class of RSMI algorithms.

A Proof of Proposition 1

In order to prove Proposition 1, let

$$\begin{bmatrix} R_{22}(\bar{t}) & \bar{z}_i(\bar{t}+1) \end{bmatrix} = R_{22}(\bar{t}+1). \tag{77}$$

Thus, from (77), it holds that

$$R_{22}(\bar{t}+1)R_{22}^T(\bar{t}+1) = R_{22}(\bar{t})R_{22}^T(\bar{t}) + \bar{z}_i(\bar{t}+1)\bar{z}_i^T(\bar{t}+1), \tag{78}$$

and note that, from Remark 1

$$Y_{t,i,j+1}\Pi_U^\perp = \underbrace{Y_{t,i,j+1}}_{\begin{bmatrix} Y_{t,i,j} & y_i(\bar{t}+1) \end{bmatrix}} - \hat{H}_i \underbrace{U_{t,i,j+1}}_{\begin{bmatrix} U_{t,i,j} & u_i(\bar{t}+1) \end{bmatrix}} = R_{22}(\bar{t}+1)Q_2(\bar{t}+1). \tag{79}$$

Then, from (79)

$$\begin{aligned}
R_{22}(\bar{t}+1)R_{22}^T(\bar{t}+1) &= \begin{bmatrix} \begin{bmatrix} Y_{t,i,j} & y_i(\bar{t}+1) \end{bmatrix} - \hat{H}_i \begin{bmatrix} U_{t,i,j} & u_i(\bar{t}+1) \end{bmatrix} \end{bmatrix} \\
& \quad \begin{bmatrix} \begin{bmatrix} Y_{t,i,j} & y_i(\bar{t}+1) \end{bmatrix} - \hat{H}_i \begin{bmatrix} U_{t,i,j} & u_i(\bar{t}+1) \end{bmatrix} \end{bmatrix}^T = \\
&= \begin{bmatrix} Y_{t,i,j} - \hat{H}_i U_{t,i,j} \end{bmatrix} \begin{bmatrix} Y_{t,i,j} - \hat{H}_i U_{t,i,j} \end{bmatrix}^T + \\
& \quad + \begin{bmatrix} y_i(\bar{t}+1) - \hat{H}_i u_i(\bar{t}+1) \end{bmatrix} \begin{bmatrix} y_i(\bar{t}+1) - \hat{H}_i u_i(\bar{t}+1) \end{bmatrix}^T.
\end{aligned} \tag{80}$$

18

Finally,
$$R_{22}(\bar{t}+1)R_{22}^T(\bar{t}+1) = R_{22}(\bar{t})R_{22}^T(\bar{t}) + z_i(\bar{t}+1)z_i^T(\bar{t}+1). \tag{81}$$

Thus, by exploiting (78) and (81), it holds that
$$\bar{z}_i(\bar{t}+1) = \pm z_i(\bar{t}+1). \tag{82}$$

B Proof of Proposition 2

First of all, consider the parameterization of observability matrix introduced in equation (45) (see Subsection 5.2.1) where Γ_{i_1} can be chosen equal to I_n without loss of generality. Since $\Gamma_i \in \mathbb{R}^{li \times 1}$ has full column rank, a thin QR factorization can be applied
$$\Gamma_i = \begin{bmatrix} I_n \\ P^T \end{bmatrix} = QR \tag{83}$$

with $Q \in \mathbb{R}^{li \times n}$ and $R \in \mathbb{R}^{n \times n}$. Then
$$\begin{bmatrix} I_n \\ P^T \end{bmatrix} R^{-1} = Q. \tag{84}$$

Since Q is an orthogonal matrix
$$Q^T Q = R^{-T} \left[I_n + PP^T \right] R^{-1} = I_n, \tag{85}$$

and therefore
$$\left[I_n + PP^T \right] = R^T R. \tag{86}$$

The above computations make it possible to express the projector $\Pi = QQ^T$ associated with the observability subspace as a function of the propagator P, as follows:
$$\Pi = QQ^T = \begin{bmatrix} I_n \\ P^T \end{bmatrix} (R^T R)^{-1} \left[I_n \ P \right]. \tag{87}$$

Then, with (85)
$$\Pi = \begin{bmatrix} I_n \\ P^T \end{bmatrix} \left[I_n + PP^T \right]^{-1} \left[I_n \ P \right] =$$
$$= \begin{bmatrix} \left[I_n + PP^T \right]^{-1} & \left[I_n + PP^T \right]^{-1} P \\ P^T \left[I_n + PP^T \right]^{-1} & P^T \left[I_n + PP^T \right]^{-1} P \end{bmatrix}. \tag{88}$$

It is now possible to express the criterion (60) as a function of the propagator P:
$$\tilde{\tilde{V}}(\Pi) = E \left\| z - \Pi z \right\|^2 =$$
$$= E \left\| \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} - \begin{bmatrix} \left[I_n + PP^T \right]^{-1} & \left[I_n + PP^T \right]^{-1} P \\ P^T \left[I_n + PP^T \right]^{-1} & P^T \left[I_n + PP^T \right]^{-1} P \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right\|^2 =$$
$$= E \left\| \begin{bmatrix} z_1 - \left[I_n + PP^T \right]^{-1} z_1 - \left[I_n + PP^T \right]^{-1} P z_2 \\ z_2 - P^T \left[I_n + PP^T \right]^{-1} z_1 - P^T \left[I_n + PP^T \right]^{-1} P z_2 \end{bmatrix} \right\|^2. \tag{89}$$

By applying the matrix inversion lemma to $\left[I_n + PP^T \right]^{-1}$, equation (89) becomes
$$\tilde{\tilde{V}}(\Pi) = E \left\| \begin{bmatrix} P \left[I_{li-n} + P^T P \right]^{-1} P^T z_1 - \left[I_n + PP^T \right]^{-1} P z_2 \\ z_2 - P^T z_1 + P^T \left[P \left[I_{li-n} + P^T P \right]^{-1} P^T z_1 - \left[I_n + PP^T \right]^{-1} P z_2 \right] \end{bmatrix} \right\|^2. \tag{90}$$

By using the above results obtained with the matrix inversion lemma, it follows that
$$P \left[P^T P + I_{li-n} \right]^{-1} = \left[I_n + PP^T \right]^{-1} P. \tag{91}$$

Then

$$\tilde{\tilde{V}}(\Pi) = E \left\| \begin{bmatrix} \left[I_n + PP^T\right]^{-1} P \left[P^T z_1 - z_2\right] \\ z_2 - P^T z_1 + P^T \left[I_n + PP^T\right]^{-1} P \left[P^T z_1 - z_2\right] \end{bmatrix} \right\|^2 = \tag{92}$$

$$= E \left\| \left[I_n + PP^T\right]^{-1} P \left[P^T z_1 - z_2\right] \right\|^2 + E \left\| \left[P^T P \left[P^T P + I_{li-n}\right]^{-1} - I_{li-n}\right] \left[P^T z_1 - z_2\right] \right\|^2.$$

The above equation shows that there is a connection between the propagator criterion (49) and Yang's cost function. In particular, it is easy to see that Yang's cost function could be interpreted as a "weighted" version of the propagator cost function.

From a practical point of view, however, it might be interesting to assess in a quantitative way the loss of performance in the estimation process which is caused by the use of the propagator method. For, note that according to equation (60) Yang's criterion can also be written as

$$\tilde{\tilde{V}}(\Pi) = tr(R_z) - tr(R_z \Pi), \tag{93}$$

and, recalling (87),

$$\tilde{\tilde{V}}(P) = tr(R_z) - tr(R_z \begin{bmatrix} I_n \\ P^T \end{bmatrix} \left[I_n + PP^T\right]^{-1} [I_n \ P]). \tag{94}$$

C   The EIVPM algorithm

$$P^T(t) = P^T(t-1) + (g(t) - P^T(t-1)\Psi(t))K(t)$$

$$g(t) = \begin{bmatrix} \hat{C}_{z_{i_2}\xi}(t-1)\xi(t) & z_{i_2}(t) \end{bmatrix}$$

$$\Lambda(t) = \begin{bmatrix} -\xi^T(t)\xi(t) & \lambda \\ \lambda & 0 \end{bmatrix}$$

$$q(t) = \hat{C}_{z_{i_1}\xi}(t-1)\xi(t)$$

$$\Psi(t) = \begin{bmatrix} q(t) & z_{i_1}(t) \end{bmatrix}$$

$$K(t) = (\Lambda(t) + \Psi^T(t)M(t-1)\Psi(t))^{-1}\Psi^T(t)M(t-1)$$

$$\hat{C}_{z_{i_1}\xi}(t) = \lambda\hat{C}_{z_{i_1}\xi}(t-1) + z_{i_1}(t)\xi^T(t)$$

$$\hat{C}_{z_{i_2}\xi}(t) = \lambda\hat{C}_{z_{i_2}\xi}(t-1) + z_{i_2}(t)\xi^T(t)$$

$$M(t) = \frac{1}{\lambda^2}(M(t-1) - M(t-1)\Psi(t)K(t))$$

$$\text{with } M(t) = (\hat{C}_{z_{i_1}\xi}(t)\hat{C}_{z_{i_1}\xi}^T(t))^{-1}$$

References

[1] M. Verhaegen and E. Deprettere. A fast, recursive MIMO state space model identification algorithm. In *Proc. of the 30th IEEE Conference on Decision and Control*, pages 1349–1354, 1991.

[2] Y. M. Cho, G. Xu, and T. Kailath. Fast recursive identification of state space models via exploitation displacement structure. *Automatica*, 30(1):45–60, January 1994.

[3] T. Gustafsson. Recursive system identification using instrumental variable subspace tracking. In *Proceedings of the 11th IFAC Symposium on System Identification*, Fukuoka, Japan, 1997.

[4] M. Lovera, T. Gustafsson, and M. Verhaegen. Recursive subspace identification of linear and non linear Wiener state-space models. *Automatica*, 36(11):1639–1650, November 2000.

[5] H. Oku and H. Kimura. Recursive 4SID algorithms using gradient type subspace tracking. *Automatica*, 38(6):1035–1043, June 2002.

[6] M. Lovera, T. Parisini, and M. Verhaegen. Fault detection: a subspace identification approach. In *Proc. of the 39th IEEE Conference on Decision and Control*, Orlando, Florida, 2001.

[7] H. Oku, G. Nijsse, M. Verhaegen, and V. Verdult. Change detection in the dynamics with recursive subspace identification. In *Proc. of the 39th IEEE Conference on Decision and Control*, Orlando, Florida, 2001.

[8] E. Hale and J. Qin. Subspace model predictive control and a case study. In *Proc. of the American Control Conference*, Anchorage, Alaska, 2002.

[9] M. Lovera and M. Verhaegen. Recursive subspace identification of linear and non linear Wiener type models. In *Proceedings of the IEEE Conference on Control Applications*, Trieste, Italy, 1998.

[10] B. Yang. Projection approximation subspace tracking. *IEEE Transactions on Signal Processing*, 43(1):95–107, January 1995.

[11] M. Lovera. Recursive subspace identification based on projector tracking. In *Proc. of the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August 2003.

[12] G. Mercère, S. Lecœuche, and C. Vasseur. A new recursive method for subsapce identification of noisy systems: EIVPM. In *Proc. of the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August 2003.

[13] W. Utschick. Tracking of signal subspace projectors. *IEEE Transactions on Signal Processing*, 50(4):769–778, 2002.

[14] J. Munier and G. Y. Delisle. Spatial analysis using new properties of the cross spectral matrix. *IEEE Transactions on Signal Processing*, 39(3):746–749, March 1991.

[15] T. McKelvey. *Identification of state space models from time and frequency data*. PhD thesis, Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden, 1995.

[16] C.T. Chou and M. Verhaegen. Subspace algorithms for the identification of multivariable dynamic error-in-variables state space models. *Automatica*, 33(10):1857–1869, 1997.

[17] M. Verhaegen and P. Dewilde. Subspace model identification, part 1: output error state-space model identification class of algorithms. *International Journal of Control*, 56(5):1187–1210, 1992.

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

[19] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input output data. *Automatica*, 30:61–74, 1994.

[20] M. Verhaegen. Subspace model identification, part 3: analysis of the ordinary output error state space model identification algorithm. *International Journal of Control*, 58(3):555–586, 1993.

[21] C. S. MacInnes and R. J. Vaccaro. Tracking directions-of-arrival with invariant subspace updating. *Signal Processing*, 50:137–150, 1996.

[22] B. Champagne and Q. C. Liu. Plane rotation-based EVD updating schemes for efficient subspace tracking. *IEEE Transactions on Signal Processing*, 46(7):1886–1900, July 1998.

[23] G. Mercère and S. Lecœuche. Identification en ligne de systèmes bruités : une nouvelle méthode récursive des sous-espaces. In *Journées Doctorales d'Automatique 2003*, Juny 2003.

[24] G. H. Golub and C. F. Van Loan. *Matrix computations*. John Hopkins University Press, Baltimore MD, 3rd edition, 1996.

[25] F. Tasker, A. Bosse, and S Fisher. Real time modal parameter estimation using subspace methods: theory. *Mechanical Systems and Signal Processing*, 12(6):797–808, 1998.

[26] G. Mercère, S. Lecœuche, and C. Vasseur. Output noise insensitive identification of systems: a new recursive state-space method. In *Proc. of the 3rd Workshop on Physics in Signal and Image Processing (PSIP)*, pages 29–32, Grenoble, France, January 2003.

[27] L. Ljung. *System identification. Theory for the user.* PTR Prentice Hall Information and System Sciences Series. T. Kailath, Series Editor, 2nd edition, 1999.

[28] S. Van Huffel and J. Vandewalle. *The total least squares problem: computational aspects and analysis.* SIAM, 1991.

[29] T. Söderström and P. Stoica. *System identification.* Prentice Hall International Series in Systems and Control Engineering, New York, 1989.

[30] B. Friedlander. The overdetermined recursive instrumental variable method. *IEEE Transactions on Automatic Control*, 4:353–356, 1984.

[31] F. D. Murnaghan. *The theory of group representations.* Dover Publications, Inc., 1938.

[32] T. Gustafsson. Instrumental variable subspace tracking using projection approximation. *IEEE Transactions on Signal Processing*, 46(3):669–681, March 1998.

[33] B. Porat and B. Friedlander. The square-root overdetermined recursive instrumental variable algorithm. *IEEE Transactions on Automatic Control*, 34(6):656–658, June 1989.