Taylor & Francis
Taylor & Francis Group

Criteria for informative experiments with subspace identification

NELSON L. C. CHUI and J. M. MACIEJOWSKI*

Cambridge University Engineering Department, Cambridge, CB2 1PZ, England

Informative experiments are identification experiments which contain sufficient information for an identification algorithm to discriminate between different models in an intended model set. In this paper, a particular set of identification algorithms, namely subspace based identification, is considered. Criteria for experiments to be informative with these methods in the deterministic setup and the combined deterministic-stochastic setup are presented. It is pointed out that if these criteria are not satisfied, interesting phenomena, in which perfect cancellations of the deterministic components and the stochastic components occur in a subspace projection, may occur. It is further shown that such cancellations can indeed be avoided under mild conditions.

1. Introduction

Subspace methods are very successful 'one-shot' (i.e. non-iterative) methods for identifying multivariable linear systems in state-space form (up to a similarity transformation) from input–output data. However, such algorithms may not necessarily yield a meaningful model in all cases. For example, suppose the input signal is identically zero. Then, any output driven by any initial state condition will not provide sufficient information about how a signal is transformed from the input to the output. In such a case, subspace algorithms will produce an incorrect model. To ensure proper use of the algorithms in this respect, a restriction on the domain of the data set is required. The study of this domain, or sets of informative experiments, is the focus of this paper.

Subspace methods have attracted a lot of attention in the system identification literature in the past few years (Moonen *et al.* 1989, Larimore 1990, Verhaegen and Dewilde 1992a, 1992b, Verhaegen 1994, Van Overschee and De Moor 1994, 1996, Viberg 1995, Deistler *et al.* 1995, Westwick and Verhaegen 1996, Chou and Verhaegen 1997).A number of issues, such as projection weighting, statistical efficiency, and numerical robustness, have been investigated extensively. However, a common implicit assumption has also been

made in these algorithms. This assumption requires an empty intersection of some particular row spaces constructed from the state sequence and the input sequence. The phenomenon in which input and state sequences violate this empty intersection assumption is known as 'rank cancellation' (De Moor 1988, p. 270). It is further shown in Jansson and Wahlberg (1998) by means of an example that even a persistently exciting input of any order may not prevent rank cancellation, if inappropriate 'block size' parameters are used in a subspace method. Rank cancellation leads to incorrect models being obtained with subspace methods, as it results in some significant 'principal components' being excluded at the initial step of subspace identification procedures; hence it usually leads to 'undermodelling'. The papers (Chui and Maciejowski 1996, Jansson and Wahlberg 1998) have addressed this issue more explicitly than the original treatment in De Moor (1988). In particular, Chui and Maciejowski (1996) addressed the deterministic case, where no stochastic excitation is present, while Jansson and Wahlberg (1998) found conditions for successful identification by subspace methods if process noise is not present or if the input signal is white noise. In this paper further analysis and derivations are presented, particularly on deterministic-stochastic cancellation in the general case, thus removing the restrictions present in Jansson and Wahlberg (1998).

We point out that the term 'deterministic' has been used (as above) in the recent 'subspace' literature to refer to that part of the process which is excited by the

*Corresponding author. Email: jmm@eng.cam.ac.uk

observed inputs, and the term 'stochastic' has been used to refer to the part excited entirely by unobserved inputs, namely process noise and measurement noise. This contrasts with the well-established meaning of 'purely deterministic' in the theory of stochastic processes. To avoid confusion, we use the term 'noise-free' rather than 'deterministic' in the rest of the paper, to refer to the modes excited by the observed inputs.

Informative experiments are identification experiments which contain sufficient information to discriminate between different models in an intended model set (Ljung 1987). However, different algorithms require different amounts of information in an experiment for successful identification. In this paper, a particular class of identification algorithms, namely subspace methods, is considered. Criteria for informative experiments with these methods in the noise-free setup and the 'combined' setup (where observed and unobserved inputs act) are presented. As will be seen, the key factor for the construction of informative experiments in noise-free identification is the emptiness of the intersection of certain input and state sequences. On the other hand, the key factor in combined identification is the emptiness of certain input and projected state sequences. Conditions for ensuring such emptiness are fully explored in this paper. *Emptiness* is not quite correct, since we usually refer to subspaces and intersections of subspaces, but it is convenient to follow the widespread convention of using 'empty' to mean that a subspace contains only the singleton $\{0\}$.

In §2 we define the system-theoretic and stochastic set-up that we will use in this paper, and we define some notations, spaces, and projections which will be used throughout, as well as some connections with Kalman filtering. In this section signals are considered to be random variables. In §3, we present two subspace algorithms for which informative experiments will be investigated. These algorithms work with real data; in this section measured signals are treated as sampled random variables. Consequently questions of large-sample properties arise, which are treated briefly. The paper is then divided into two parts, in which §4 and 5 deal with the noise-free case, whereas §7 and 8 deal with the combined case. Section 4 contains a definition of 'sufficiently rich' input sequences, which is used throughout the rest of the paper. To facilitate the study of the combined case, §6 presents two useful subsystem decompositions; one decomposes a system into noise-free and stochastic subsystems which are uncorrelated with each other, but the stochastic part is then non-stationary, while the other decomposes the system into a Kalman filter and a complementary error subsystem. Most of the proofs have been placed in the Appendix.

An earlier version of this paper appeared as Chui and Maciejowski (1998a).

Throughout this paper, the set of real numbers is denoted by $\mathbb{R}$. The set of integers and non-negative integers are denoted by $\mathbb{Z}$ and $\mathbb{Z}_+$, respectively. The general Moore–Penrose inverse is written as † while the Hermitian as *. Denote by $+$, $\oplus$ and $\cap$ the sum, the direct sum and the intersection of two vector spaces. The notation $^{\perp}$ denotes the orthogonal complement of a subspace with respect to the predefined ambient space.

2. System description

Consider a standard linear system with the state-space equations

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

$$y(t) = Cx(t) + Du(t) + v(t), \tag{1b}$$

where the system matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$. Here, u and y are the input and the output signal, respectively, whereas w and v are the process and the output (measurement) noise, respectively.

Let Ω be a sample space, $\mathcal{F}$ be a σ-field of subsets of Ω, and $\mathbf{P}$ be a probability measure on $(\Omega, \mathcal{F})$. A real random variable ψ is a function $\psi : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : \psi(\omega) \leq c\} \in \mathcal{F}$ for each $c \in \mathbb{R}$. Denote by $\mathcal{L}_2(\Omega, \mathcal{F}, \mathbf{P})$ the space of all real Lebesgue square integrable random variables on $(\Omega, \mathcal{F}, \mathbf{P})$

$$\mathcal{L}_2(\Omega, \mathcal{F}, \mathbf{P}) := \left\{ \psi : \int_{\Omega} \psi^2(\omega) \, d\mathbf{P}(\omega) < \infty \right\}.$$

In this paper, we consider the setup in which, for any time t, the signals $x(t)$, $u(t)$, $y(t)$, $w(t)$ and $v(t)$ are $\mathcal{L}_2(\Omega, \mathcal{F}, \mathbf{P})$-valued vectors. This is to say $x(t)$, for instance, is an element of $\mathcal{L}_2^n(\Omega, \mathcal{F}, \mathbf{P})$. Denote by $\mathbf{E}$ the usual expectation operator. Moreover, let Z^N be a finite sequence of input-output signals

$$Z^N = \big(y(0), u(0), y(1), u(1), \ldots, y(N-1), u(N-1) \big). \tag{2}$$

We emphasise again that signals are considered to be random variables here, so that Z^N is a finite sequence of random variable-valued vectors.

The linear systems of concern in this paper are those comprised of two subsystems, a subsystem with observable inputs and a subsystem with unobservable inputs. The former subsystem, referred to as a noise-free subsystem (with a superscript d as mnemonic for 'deterministic'), describes the dynamics due to the input u

$$x^d(t+1) = Ax^d(t) + Bu(t), \tag{3a}$$

$$y^d(t) = Cx^d(t) + Du(t); \tag{3b}$$

whereas the latter subsystem, referred to as a stationary stochastic subsystem, describes the stationary stochastic process due to the process and output noises w and v

$$x^s(t+1) = Ax^s(t) + w(t), \tag{4a}$$
$$y^s(t) = Cx^s(t) + v(t). \tag{4b}$$

The state and the output of the combined system are then the sum of those of the subsystems as shown in the following

$$x(t) = x^d(t) + x^s(t),$$
$$y(t) = y^d(t) + y^s(t).$$

The set of systems described above has been studied extensively in the system identification literature.

2.1. A Stationary stochastic subsystem

Stationary stochastic processes are considered as follows. The noises w and v are white-noise processes with the property

$$\mathbf{E}\left[\begin{pmatrix} w(t) \\ v(t) \end{pmatrix} \begin{pmatrix} w(\tau) \\ v(\tau) \end{pmatrix}^* \right] =: \begin{pmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{pmatrix} \delta_{t\tau}, \tag{5}$$

for all $t, \tau \in \mathbb{Z}_+$, where δ denotes the Kronecker delta. Here, Σ^w, Σ^v, Σ^{wv}, and Σ^{vw} are real constant matrices with obvious dimensions. Moreover, define the state correlation matrix Σ^s, the output correlation matrix Λ_i, and the cross input–output correlation matrix G as

$$\Sigma^s := \mathbf{E}\,(x^s(t)\,[x^s(t)]^*),$$
$$\Lambda_i := \mathbf{E}\,(y^s(t+i)\,[y^s(t)]^*),$$
$$G := \mathbf{E}\,(x^s(t+1)\,[y^s(t)]^*),$$

for $t \geq i$, $i \in \mathbb{Z}$ and $t \in \mathbb{Z}_+$. Note that since the stochastic subsystem of equations (4) is assumed stationary, the above matrices are constant matrices independent of t. Consequently, the following identities can be derived

$$\Sigma^s = A\Sigma^s A^* + \Sigma^w,$$
$$\Lambda_0 = C\Sigma^s C^* + \Sigma^v,$$
$$G = A\Sigma^s C^* + \Sigma^{wv}.$$

Furthermore, for $i \neq 0$, Λ_i is equal to

$$\Lambda_i = CA^{i-1}G,$$
$$\Lambda_{-i} = G^*(A^{i-1})^* C^*.$$

Finally, the system defined in equations (1) is assumed to be minimal in the sense that the pair (A, C) is observable and the pair $(A, [B\ (\Sigma^w)^{1/2}])$ is controllable.

2.2. Projection spaces

In subspace identification methods, observations at each time instant are 'lifted' up into a sequence of observations. These sequences are then stacked together, forming a block Hankel matrix as in the following. Given an input–output signal pair Z^N, denote by U the block Hankel matrix constructed from the input signal u

$$U := \begin{bmatrix} u(0) & u(1) & \cdots & u(N-h) \\ u(1) & u(2) & \cdots & u(N-h+1) \\ \ddots & \ddots & \ddots & \ddots \\ u(h-1) & u(h) & \cdots & u(N-1) \end{bmatrix}. \tag{6}$$

Analogously, denote by X, Y, W, and V the corresponding block Hankel matrix constructed from x, y, w and v. Furthermore, denote by additional subscripts i, j the submatrix starting from the ith block row to the jth block row, as

$$U_{i,j} := \begin{bmatrix} u(i) & u(i+1) & \iddots & u(N-h+i) \\ u(i+1) & u(i+2) & \iddots & u(N-h+i+1) \\ \ddots & \ddots & \ddots & \ddots \\ u(j) & u(j+1) & \cdots & u(N-h+j) \end{bmatrix}.$$

To indicate one block row, that is for $i = j$, the submatrix will be written simply as U_i. Moreover, let $q = N - h + 1$ be the number of columns of the block Hankel matrix U, or the length of each individual sequence U_i. With this setup, the system equations can be extended to the following in a straightforward manner

$$X_{t+1} = AX_t + BU_t + W_t, \tag{7a}$$
$$Y_t = CX_t + DU_t + V_t. \tag{7b}$$

Again, the noise-free subsystem has the state-space equations

$$X_{t+1}^d = AX_t^d + BU_t,$$
$$Y_t^d = CX_t^d + DU_t;$$

and the stochastic subsystem has the state-space equations

$$X_{t+1}^s = AX_t^s + W_t,$$
$$Y_t^s = CX_t^s + V_t.$$

Define the noise-free controllability matrix $\mathcal{C}_i^d$, the stochastic controllability matrix $\mathcal{C}_i^s$, and the 'full'

controllability matrix $\mathcal{C}_i^w$ as

$$\mathcal{C}_i^d := [A^{i-1}B \quad A^{i-2}B \quad \cdots \quad B],$$
$$\mathcal{C}_i^s := [A^{i-1}G \quad A^{i-2}G \quad \cdots \quad G],$$
$$\mathcal{C}_i^w := [A^{i-1}I \quad A^{i-2}I \quad \cdots \quad I].$$

Furthermore, define the observability matrix $\mathcal{O}_i$, and block-Toeplitz matrices $\mathcal{T}_i^d$ and $\mathcal{T}_i^w$ as

$$\mathcal{O}_i := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{bmatrix}, \quad \mathcal{T}_i^d := \begin{bmatrix} D & & & 0 \\ CB & D & & \\ \vdots & \ddots & \ddots & \\ CA^{i-2}B & \cdots & CB & D \end{bmatrix},$$

$$\mathcal{T}_i^w := \begin{bmatrix} 0 & & & 0 \\ C & 0 & & \\ \vdots & \ddots & \ddots & \\ CA^{i-2} & \cdots & C & 0 \end{bmatrix}.$$

With these matrices and the block Hankel data matrices defined above, it is easy to obtain the block system equations

$$X_t = A^t X_0 + \mathcal{C}_t^d U_{0,t-1} + \mathcal{C}_t^w W_{0,t-1}, \tag{8a}$$

$$Y_{0,t-1} = \mathcal{O}_t X_0 + \mathcal{T}_t^d U_{0,t-1} + \mathcal{T}_t^w W_{0,t-1} + V_{0,t-1}. \tag{8b}$$

Here all matrices have h block rows, except the state signal X_0 which has only one block row. The first equation is usually referred to as the state sequence equation, and the second one is referred to as the block data equation.

Spaces spanned by subsequences of signals are very prominent in this paper. These spaces can be envisaged as the row spaces of the block Hankel matrices defined above. To facilitate the analysis, denote by $\mathcal{U}$ the space spanned by all the rows of U. That is,

$$\mathcal{U} := \text{span}\{\, \alpha^* U \mid \alpha \in \mathbb{R}^{hm} \,\}.$$

Similar notation will be used to represent the row spaces spanned by all other block Hankel matrices. Furthermore, let Π be the orthogonal projection operator. In this paper, we will use the shorthand $\Pi_{\mathcal{R}}Q$ to represent the orthogonal projection of each row of Q onto the space $\mathcal{R}$. For instance, the expression $\Pi_{\mathcal{U}}X_0$ denotes

$$\Pi_{\mathcal{U}}X_0 := \begin{bmatrix} \Pi_{\mathcal{U}}X_{0;1} \\ \vdots \\ \Pi_{\mathcal{U}}X_{0;n} \end{bmatrix}, \quad \text{where } X_0 = \begin{bmatrix} X_{0;1} \\ \vdots \\ X_{0;n} \end{bmatrix}.$$

This projection can be computed as $\mathbf{E}(X_0 U^*)\mathbf{E}(U U^*)^\dagger U$. Finally, it can be seen without difficulty that $\mathbf{E}(\mathcal{O}_h X_0 U^*)\mathbf{E}(U U^*)^\dagger U$ can also be written as $\mathcal{O}_h \mathbf{E}(X_0 U^*)\mathbf{E}(U U^*)^\dagger U$, or equivalently, $\Pi_{\mathcal{U}}\mathcal{O}_h X_0 = \mathcal{O}_h \Pi_{\mathcal{U}}X_0$.

It is recalled that definitions such as (6) refer to finite numbers of random variables, not to finite segments of sample paths of these variables, and hence a definition in terms of the expectation operator is appropriate here. We will consider the treatment of real data later, in §3.

2.3. Kalman filtering

In this section, a link between the Kalman filters for the two systems, equations (7) and (8), is presented. Note that the systems considered in this paper are not restricted to fully stochastically excited systems. That is to say, some states or outputs may not be excited by unobservable noises w and v—they may be 'noise-free'. Because of this possibility, a transformation which segregates noise-free and stochastic modes is provided in the following proposition.

Proposition 1 (Chui 1997, Chui and Maciejowski 1998b)**:** *For any observable state-space realization of equations* (7), *there always exist an invertible T_s and a unitary T_o such that*

$$\bar{X}_t := T_s X_t = \begin{bmatrix} T_{s;1} \\ T_{s;2} \\ T_{s;3} \end{bmatrix} \quad X_t =: \begin{bmatrix} \bar{X}_{t;1} \\ \bar{X}_{t;2} \\ \bar{X}_{t;3} \end{bmatrix},$$

$$\bar{Y}_t := T_o Y_t = \begin{bmatrix} T_{o;1} \\ T_{o;2} \\ T_{o;3} \end{bmatrix} \quad Y_t =: \begin{bmatrix} \bar{Y}_{t;1} \\ \bar{Y}_{t;2} \\ \bar{Y}_{t;3} \end{bmatrix}, \tag{9a}$$

and

$$\bar{\Sigma}^s := T_s \Sigma^s T_s^* = \begin{bmatrix} \bar{\Sigma}_{11}^s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\bar{\Sigma}^v := T_o \Sigma^v T_o^* = \begin{bmatrix} \bar{\Sigma}_{11}^v & \bar{\Sigma}_{12}^v & 0 \\ \bar{\Sigma}_{21}^v & \bar{\Sigma}_{22}^v & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{9b}$$

and

$$\bar{A} := T_s A T_s^{-1} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix},$$

$$\bar{C} := T_o C T_s^{-1} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ 0 & \bar{C}_{22} & \bar{C}_{23} \\ 0 & 0 & \bar{C}_{33} \end{bmatrix}, \tag{9c}$$

where

$$\bar{\Sigma}_{11}^s > 0, \quad \bar{\Sigma}_{22}^v > 0, \quad \bar{C}_{11}\bar{C}_{11}^* > 0,$$
$$(\bar{C}_{11}, \bar{A}_{11}) \quad and \quad (\bar{C}_{33}, \bar{A}_{33}) \text{ observable.} \qquad (9d)$$

Note that $\bar{X}_{t;1} := T_{s;1}X_t$ corresponds to the stochastically excited modes; $\bar{X}_{t;2} := T_{s;2}X_t$ corresponds to the output disturbance modes; and $\bar{X}_{t;3} := T_{s;3}X_t$ corresponds to modes that are 'noise-free'. According to this transformation, the reduced stochastic system, in the state and output coordinates defined in equations (9), has the state-space equations

$$\bar{X}_{t+1;1}^s = \bar{A}_{11} \bar{X}_{t;1}^s + \bar{W}_{t;1}, \qquad (10a)$$

$$\begin{bmatrix} \bar{Y}_{t;1}^s \\ \bar{Y}_{t;2}^s \end{bmatrix} = \begin{bmatrix} \bar{C}_{11} \\ 0 \end{bmatrix} \bar{X}_{t;1}^s + \begin{bmatrix} \bar{V}_{t;1} \\ \bar{V}_{t;2} \end{bmatrix}. \qquad (10b)$$

where X^s and Y^s correspond to x^s and y^s, respectively—see (4). Furthermore, this reduced system is assumed to be strictly positive real in order to avoid introducing further complication into this paper.

Now, applying standard Kalman filtering to equations (7), it can be shown that the corresponding Kalman filter has the form

$$\hat{X}_t = A\hat{X}_{t-1} + BU_{t-1}$$
$$+ K_{t-1}(Y_{t-1} - C\hat{X}_{t-1} - DU_{t-1}), \qquad (11a)$$

$$K_{t-1} = (G + A\tilde{P}_{t-1}C^*)(\Lambda_0 + C\tilde{P}_{t-1}C^*)^\dagger, \qquad (11b)$$

$$\tilde{P}_t = A\tilde{P}_{t-1}A^* - (G + A\tilde{P}_{t-1}C^*)$$
$$\times (\Lambda_0 + C\tilde{P}_{t-1}C^*)^\dagger(G + A\tilde{P}_{t-1}C^*)^*, \qquad (11c)$$

where

$$\tilde{P}_0 + \Sigma^s := \frac{1}{q}\mathbf{E}([X_0 - \hat{X}_0][X_0 - \hat{X}_0]^*). \qquad (12)$$

Here, q is the length of the sequences X_0 and $\hat{X}_0$. Note that the general Moore–Penrose inverse is used in these equations, so that they apply to cases where some state coordinates are 'noise-free'. However, in order for the above formulae to hold, certain restrictions apply to the state estimate $\hat{X}_0$, as shown in Chui (1997), Chui and Maciejowski (1998b). Before looking at the block system equations, first define the following

Toeplitz matrix L_i as

$$L_i := \begin{bmatrix} \Lambda_0 & \Lambda_{-1} & \cdots & \Lambda_{1-i} \\ \Lambda_1 & \Lambda_0 & \ddots & \Lambda_{2-i} \\ \vdots & \ddots & \ddots & \vdots \\ \Lambda_{i-1} & \Lambda_{i-2} & \cdots & \Lambda_0 \end{bmatrix}.$$

Using the same Kalman filtering technique on equations (8), it can be shown that the corresponding non-steady state Kalman filter has the form

$$\hat{X}_t = A^t\hat{X}_0 + \mathcal{C}_t^d U_{0,t-1}$$
$$+ \mathcal{K}_t(Y_{0,t-1} - \mathcal{O}_t\hat{X}_0 - \mathcal{T}_t^d U_{0,t-1}), \qquad (13a)$$

$$\mathcal{K}_t := (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^\dagger, \qquad (13b)$$

$$\tilde{P}_t := A^t\tilde{P}_0(A^t)^* - (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)$$
$$\times (L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^\dagger(\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)^*. \qquad (13c)$$

Van Overschee and De Moor (1994) showed that the Kalman filter defined in equation (13) is actually a *t*-step ahead Kalman filter generated from equation (11). Extension to the partially stochastic case has also been given in Chui (1997), Chui and Maciejowski (1998b). This is summarized in the following theorem.

Theorem 1 (Van Overschee and De Moor 1994): *Suppose that $\tilde{P}_0$ satisfies equation* (12) *for some initial state estimate $\hat{X}_0$. Then, $\hat{X}_t$ generated recursively from equations* (11) *is equal to $\hat{X}_t$ computed directly from equations* (13), *for any $t \in \mathbb{Z}_+$.*

3. Subspace algorithms

Up to now, an ideal probabilistic setup has been considered. To deal with real data, the technique of replacing the expectation operator $\mathbf{E}$ by an 'averaging' operator $\bar{\mathbf{E}}$ has been employed in the literature (Van Overschee and De Moor 1994, Jansson and Wahlberg 1998). One straightforward setup is to have an infinite number of experiments. For instance, let real sequences $\{y^1(i), y^2(i), y^3(i), \ldots\}$ and $\{u^1(i), u^2(i), u^3(i), \ldots\}$ be realizations of $y(i)$ and $u(i)$ taken from different experiments. Then,

$$\bar{\mathbf{E}}[y(i)u^*(i)] = \lim_{T \to \infty} \frac{1}{T}\sum_{j=1}^T y^j(i)u^{j*}(i).$$

Another setup is to have a long time series from one experiment. Then, $\bar{\mathbf{E}}$ is defined to be averaging over an infinitely long experiment. In this case, signals are assumed to be ergodic or quasi-stationary (Ljung 1987, Jansson and Wahlberg 1998). For instance, let real sequences $\{y^1(0), y^1(1), y^1(2), \ldots\}$ and $\{u^1(0), u^1(0), u^1(2), \ldots\}$ be realizations of y and u obtained from one experiment. Then,

$$\bar{\mathbf{E}}[y(i)u^*(i)] = \lim_{T \to \infty} \frac{1}{T} \sum_{j=0}^{T-1} y^1(i+jq)u^{1*}(i+jq).$$

This is the setup which we will assume. Since $\bar{\mathbf{E}}[.] = \mathbf{E}[.]$ almost surely, we shall replace $\mathbf{E}[.]$ by $\bar{\mathbf{E}}_T[.]$ in the algorithms, where

$$\bar{\mathbf{E}}_T[y(i)u^*(i)] = \frac{1}{T} \sum_{j=0}^{T-1} y^1(i+jq)u^{1*}(i+jq). \quad (14)$$

All our results will then hold almost surely as $T \to \infty$.

Although many versions of subspace algorithms have been developed, they all share a similar methodology, and most of them have been shown to be equivalent, up to a choice of weighting functions in the observability/state sequence decomposition (Van Overschee and De Moor 1995). In this section we present two generic subspace algorithms, for which informative experiments will be investigated. In particular, the first algorithm (Algorithm 3.1) is analogous to the ordinary MOESP algorithm (Verhaegen 1993) in which no instrumental variable is used. The second algorithm (Algorithm 3.2) is analogous to the MOESP PO scheme (Verhaegen 1994), which employs instrumental variable techniques. The only difference between the algorithms presented here and the MOESP algorithms is that the MOESP algorithms are formulated in RQ factorisations for computational consideration, instead of direct operations on U and Y. (Also see Chui and Maciejowski (1998a) for further exposition.) Finally, note that by exploiting their similarity, results developed in this paper are readily extended to other algorithms which have been proposed.

A one-block subspace algorithm
First of all, an algorithm (Algorithm 3.1) which can identify systems in a noise-free environment is presented. Its requirements for informative experiments will be discussed in § 5. Note that singular value decomposition is involved in the algorithm: $\mathbf{\Pi}_{\mathcal{U}^\perp} Y = \Gamma \Sigma \Omega$. Since Y is a random variable, this decomposition should be interpreted in the following way: Γ is real-valued unitary, Σ is non-increasing nonnegative diagonal, and Ω is random-variable-valued with $\mathbf{E}[\Omega\Omega^*] = I$ (See Chui

Table 1. A schematic overview of the one-block subspace algorithm.

Algorithm 3.1. One-block subspace algorithm

1. *Construct U and Y as in equation* (6).
2. *Compute the SVD of $\mathbf{\Pi}_{\mathcal{U}^\perp} Y$ as* (*with obvious partitions*)

$$\mathbf{\Pi}_{\mathcal{U}^\perp} Y = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}, \quad where \ \Gamma_1 = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_h \end{bmatrix}. \quad (15)$$

3. *Compute A and C from*

$$\begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{h-1} \end{bmatrix} A = \begin{bmatrix} \gamma_2 \\ \vdots \\ \gamma_h \end{bmatrix} \quad and \quad C = \gamma_1.$$

4. *Construct $\mathcal{O}_h^\perp$ such that $\mathcal{O}_h^\perp \mathcal{O}_h = 0$. Compute $S_0, \ldots, S_{h-1}$ from*

$$\mathcal{O}_h^\perp Y := S_0 U_0 + \cdots + S_{h-1} U_{h-1}. \quad (16)$$

5. *Determine B and D from*

$$\begin{bmatrix} S_0 \\ S_1 \\ \vdots \\ S_{h-1} \end{bmatrix} = \begin{bmatrix} P_0 & \cdots & P_{h-2} & I \\ P_1 & \cdots & I & 0 \\ \vdots & \ddots & \ddots & \vdots \\ I & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \mathcal{O}_{h-1} \end{bmatrix} \begin{bmatrix} D \\ B \end{bmatrix}, \quad (17)$$

where

$$\begin{bmatrix} P_0 & \cdots P_{h-2} & I \end{bmatrix} := \mathcal{O}_h^\perp.$$

6. *Obtain the set of all state-space models*
$$\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \text{ invertible}\}.$$

(1997) for a procedure of computing this SVD). The algorithm is defined in table 1.

A two-block subspace algorithm
In the following, another subspace algorithm (Algorithm 3.2) is presented. This algorithm can give consistent estimates in the presence of process and output noises under mild conditions, though noise cancellation properties will not be discussed here. Requirements for experiments to be informative with this algorithm will be presented for the noise-free case in § 5 and the combined case in § 8.

In the following, we define the 'past' and 'future' data blocks. Here, the definition has a more general structure than usually assumed, in which the number of block rows in the 'past' data and the 'future' data can be different. This unequal splitting of data blocks has been addressed in Wahlberg and Jansson (1994),

Jansson and Wahlberg (1995) and also plays an important role in Markov parameter estimation, as discussed in Chui and Maciejowski (1998c).

$$
\begin{bmatrix} U_p \\ ---- \\ U_f \end{bmatrix}
$$

$$
:= \begin{bmatrix} u(0) & u(1) & \cdots & u(N-k-l) \\ \ddots & \ddots & \ddots & \ddots \\ u(k-1) & u(k) & \cdots & u(N-l-1) \\ ----- & ----------- \\ u(k) & u(k+1) & \cdots & u(N-l) \\ \ddots & \ddots & \ddots & \ddots \\ u(k+l-1) & u(k+l) & \cdots & u(N-1) \end{bmatrix}.
$$

(18)

Here, U_p has k block rows while U_f has l block rows. Y_p and Y_f are defined similarly. On the other hand, we define two vectors, each consisting of one block

Table 2.   A schematic overview of the two-block subspace algorithm.

Algorithm 3.2.   Two-block subspace algorithm

1. *Construct U_p, U_f and Y_p, Y_f as in equation* (18).
2. *Compute the SVD of $\mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_{\mathcal{S}} Y_f$ for $\mathcal{S} = Y_p + \mathcal{U}_p + \mathcal{U}_f$ as*

$$
\mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_{\mathcal{S}} Y_f = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}, \quad where \quad \Gamma_1 = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_l \end{bmatrix}.
$$

(19)

3. *Compute A and C from*

$$
\begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{l-1} \end{bmatrix} A = \begin{bmatrix} \gamma_2 \\ \vdots \\ \gamma_l \end{bmatrix} \quad and \quad C = \gamma_1.
$$

4. *Construct $\mathcal{O}_l^\perp$ such that $\mathcal{O}_l^\perp \mathcal{O}_l = 0$. Compute $S_0, \ldots, S_{l-1}$ from*

$$
\mathcal{O}_l^\perp \mathbf{\Pi}_{\mathcal{S}} Y_f = S_0 U_k + \cdots + S_{l-1} U_{k+l-1}.
$$

5. *Determine B and D from*

$$
\begin{bmatrix} S_0 \\ S_1 \\ \vdots \\ S_{l-1} \end{bmatrix} = \begin{bmatrix} P_0 & \cdots & P_{l-2} & I \\ P_1 & \cdots & I & 0 \\ \vdots & & & \vdots \\ I & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \mathcal{O}_{l-1} \end{bmatrix} \begin{bmatrix} D \\ B \end{bmatrix},
$$

where

$$
\begin{bmatrix} P_0 & \cdots P_{l-2} & I \end{bmatrix} := \mathcal{O}_l^\perp.
$$

6. *Obtain the set of all state-space models*

$$
\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \text{ invertible}\}.
$$

row, $X_p := X_0$ and $X_f := X_k$. The suffices p and f are mnemonic, representing 'past' and 'future', respectively. The algorithm is defined in table 2.

Remark 1: When dealing with finite data samples, the zero singular values shown in equation (19) will be replaced by non-zero values in general, which will converge to zero as $T \to \infty$. Model reduction is required to bring the model down to a reasonable state dimension, and this is done by the usual expedient of replacing sufficiently small singular values by zero values. We assume that the decision rule for making these replacements is such that, as the sample size increases, the number of retained non-zero singular values converges to the true state dimension of the system, unless some perfect noise-free/stochastic cancellation occurs. Such a cancellation is investigated further in the following sections.

4. Sufficiently rich inputs and state sequences

The concepts of persistent excitation and sufficient richness have been used extensively in adaptive control and identification (Moore 1983, Mareels 1984, Bai and Sastry 1985, Ljung 1987, Stoica and Söderström 1989, and references therein). The definition of sufficient richness of a signal is closely related to its block Hankel matrix as defined in equation (6), whose size is dependent on the parameter h. For this purpose, denote by $U(h)$ the block Hankel matrix U with h block rows.

Definition 1: A signal u is said to be sufficiently rich (S.R.) of order h, if h ($\leq N$) is the greatest integer such that $\mathbf{E}[U(h)U^*(h)] > 0$.

This section studies the relationship between input and state sequences. Define an annihilating polynomial of A to be a polynomial $p(z)$ such that $p(A) = 0$. It is well known that all the annihilating polynomials of A form a principal ideal. The minimal polynomial $p_{\min}(z)$, which is the monic generator of this ideal, is called the A-annihilator.

The following two theorems are closely related. The first one covers the case where $\Sigma^w = 0$, and will be used to prove Theorems 1 and 2 in the noise-free case, whereas the second one covers a more general case, and will be used to prove Theorem 12 for the combined case. Here, the one-block configuration (equation 6) is considered; and hence h refers to the number of rows in the block Hankel matrices.

Theorem 2: *Given an input signal u and some state signal x satisfying equation* (1) *with*

1. $\Sigma^w = 0$;
2. (A, B) *reachable*;

3. *r no smaller than the degree of the A-annihilator*;
4. *the input signal u S.R. of order $h + r$.*

Then, $\dim\{\mathbf{\Pi}_{\mathcal{U}^\perp}\mathcal{X}_0\} = n$.

The proof of the above theorem can be found in Appendix A. The next theorem covers the case where $(A, [B \ (\Sigma^w)^{1/2}])$ is reachable; its proof can also be found in Appendix A.

Theorem 3: *Given an input signal u and some state signal x satisfying equations* (1) *with*

1. $(A, [B \ (\Sigma^w)^{1/2}])$ *reachable*;
2. *r no smaller than the degree of the A-annihilator*;
3. *the input signal u S.R. of order $h + r$.*

Then, $\dim\{\mathbf{\Pi}_{\mathcal{U}^\perp}\mathcal{X}_0\} = n$.

The above theorems imply that if u is S.R. of order no smaller than its number of block rows plus the degree of the *A*-annihilator, then $\mathcal{U}$ and $\mathcal{X}_0$ will have an empty intersection.

5. Requirements for informative experiments: noise-free case

In this section, sufficient conditions for informative experiments for subspace methods are presented for the two subspace algorithms presented in §3. It should be noted that similar studies using different approaches have also been carried out in Chui and Maciejowski (1996), Jansson and Wahlberg (1998), while the derivations in this paper lead to slightly tighter results. To conduct the analysis, we shall confine the domain of data sets to all data sets which are generated by *n*th order linear systems with some initial condition $x(0)$. Obviously, for noise-free identification we have $\Sigma^s = \Sigma^v = 0$. Moreover, we recall that the observability index is defined to be the smallest value of h such that $\mathcal{O}_h$ is injective (Kailath 1980).

The exact input conditions for Algorithm 3.1 are stated in the following theorem. It can be seen that for the noise-free case, the excitation of the input signal determines the informativeness of an experiment.

Theorem 4: *Let Z^N be an input–output signal pair of a minimal system* (1), *with*

1. $\Sigma^s = \Sigma^v = 0$;
2. *the degree of the A-annihilator no greater than r*;
3. *the observability index of the system no greater than $h - 1$*;
4. *u S.R. of order $h + r$.*

Then, *Algorithm* 3.1 *identifies the system* $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \text{ invertible}\}$.

The proof of this theorem can be found in Appendix A.

Remark 2: Consider an *n*-order single-input single-output (SISO) linear system. It is known that such a system can be represented by an *n*th order proper rational function, or an *n*-order ARMAX model. Thus, it is easy to see that a SISO system has $2n + 1$ parameters. As shown in Kim and Cain (1983), an *n*th order rational function can be determined by any input signal which is S.R. of order $2n + 1$. Accordingly, it is possible to identify a multi-input multi-output (MIMO) linear system by first identifying all its input-output channels using an input signal S.R. of order $2n + 1$. The MIMO system can then be recovered by collecting all the channels together. This identification procedure was also applied in McKelvey and Akçay (1995). From Theorem 4 we see that S.R. of order $2n + 1$ is not the smallest requirement which uniquely specifies an *n*th order MIMO system. In general, S.R. of order $h + r$ will be sufficient; both $h - 1$ and r may be less than n.

The exact input conditions for Algorithm 3.2 are stated in the following theorem. Again, the excitation of the input signal is the key to getting an informative experiment. However, since an instrumental variable $\mathcal{S}$ is used, the richness of the input excitation has to be increased to account for the extra past data block. It is also noteworthy that the number of block rows in the past data block must be no smaller than the observability index, while the number of block rows in the future data block must be greater than the observability index.

Theorem 5: *Let Z^N be an input–output signal pair of a minimal system* (1), *with*

1. $\Sigma^s = \Sigma^v = 0$;
2. *the degree of the A-annihilator no greater than r*;
3. *the observability index of the system no greater than* $\min\{k, l - 1\}$;
4. *u S.R. of order $k + l + r$.*

Then, *Algorithm* 3.2 *identifies the system* $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \text{ invertible}\}$.

Proof: Notice that if $\mathbf{\Pi}_{\mathcal{S}}Y_f$ and U_f in Algorithm 3.2 are replaced by Y and U, the algorithm will then be equivalent to Algorithm 3.1. Thus, in this proof it is shown that conditions 1–4 in this theorem do imply those in Theorem 4 and $\mathbf{\Pi}_{\mathcal{S}}Y_f = Y_f$.

First of all, conditions 1 and 2 in this theorem are identical to those in Theorem 4. Moreover, condition 3 in this theorem implies that in Theorem 4. For condition 4, one can check that U_f is S.R. of order $l + r$ by a straightforward verification of Definition 1. Now, it remains to prove $\mathbf{\Pi}_{\mathcal{S}}Y_f = Y_f$, or equivalently, $\mathcal{Y}_f \subset \mathcal{S}$.

From equation (8b), it can be seen $\mathcal{Y}_f \subset \mathcal{X}_f + \mathcal{U}_f$. Note that $X_f = A^k X_p + \mathcal{C}_k^d U_p$, which in turn implies

$\mathcal{X}_f \subset \mathcal{X}_p + \mathcal{U}_p$. Combining the two inclusions gives $\mathcal{Y}_f \subset \mathcal{X}_p + \mathcal{U}_p + \mathcal{U}_f$. Noticing $Y_p = \mathcal{O}_k X_p + \mathcal{T}_k^d U_p$ and $\mathcal{O}_k X_p = Y_p - \mathcal{T}_k^d U_p$ gives

$$\mathcal{Y}_p \subset \mathcal{X}_p + \mathcal{U}_p \quad \text{and} \quad \mathcal{X}_p \subset \mathcal{Y}_p + \mathcal{U}_p,$$

where the latter inclusion stems from the fact that $\mathcal{O}_k$ is injective. Adding $\mathcal{U}_p$ to both sides gives

$$\mathcal{Y}_p + \mathcal{U}_p = \mathcal{X}_p + \mathcal{U}_p. \qquad (20)$$

That is, $\mathcal{Y}_f \subset \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f = \mathcal{S}$, which completes the proof. $\qquad\square$

6. Decomposition into noise-free/stochastic subsystems

In this section we introduce two system decompositions. These decompositions provide two approaches for the analysis of the spaces of input and projected state sequences in the next section.

Given U_t and Y_t for $0 \le t \le h-1$, decompose X_0 into three components: $\hat{X}_0$, $\tilde{X}_0$, and X_0^s. Here, $\hat{X}_0$ and $\tilde{X}_0$ are defined as

$$\hat{X}_0 := \mathbf{\Pi}_{\mathcal{Z}^d} X_0 \quad \text{and} \quad \tilde{X}_0 := X_0^d - \hat{X}_0, \qquad (21)$$

where

$$\mathcal{Z}^d := \mathcal{U}_{0,h-1} + (\mathcal{Y}_{0,h-1}^d \cap \mathcal{Y}_{0,h-1}).$$

Recall that $\mathcal{Y}_t^d$ is obtained from the noise-free subsystem, which is uncorrelated with $\mathcal{W}_t$ and $\mathcal{V}_t$. Therefore, the intersection $\mathcal{Y}_{0,h-1}^d \cap \mathcal{Y}_{0,h-1}$ produces a space which is orthogonal to $\mathcal{W}_t$ and $\mathcal{V}_t$. The following then hold

$$\hat{\mathcal{X}}_0, \tilde{\mathcal{X}}_0, \mathcal{X}_0^s, \mathcal{W}_t \text{ and } \mathcal{V}_t \text{ are mutually orthogonal;}$$
$$(22a)$$

$$\tilde{\mathcal{X}}_0 \text{ and } \mathcal{X}_0^s \text{ are orthogonal to } \mathcal{U}_t \text{ and } (\mathcal{Y}_t^d \cap \mathcal{Y}_t). \quad (22b)$$

This decomposition of the initial state gives rise to two different system decompositions, as will be discussed in the following subsections.

6.1. *Decomposition with a non-stationary stochastic subsystem*

In this subsection, $\tilde{X}_0$, which originally belongs to the noise-free initial condition $X_0^d = \hat{X}_0 + \tilde{X}_0$, will be treated as part of the initial condition in the stochastic subsystem, creating a non-stationary stochastic subsystem with X_0^s.

Define the initial state of the non-stationary stochastic subsystem to be $X_0^{ns} := X_0^s + \tilde{X}_0$, and the initial state of the complementary noise-free subsystem to be $X_0^{nd} := \hat{X}_0$. Note that the additional superscript n signifies the non-stationary property of the stochastic subsystem. Now, define X_t^{ns} and Y_t^{ns} recursively from the following non-stationary stochastic subsystem

$$X_{t+1}^{ns} = AX_t^{ns} + W_t,$$
$$Y_t^{ns} = CX_t^{ns} + V_t;$$

and define X_t^{nd} and Y_t^{nd} recursively from the following complementary noise-free subsystem

$$X_{t+1}^{nd} = AX_t^{nd} + BU_t,$$
$$Y_t^{nd} = CX_t^{nd} + DU_t.$$

Note that it can be verified without difficulty that

$$X_t = X_t^{nd} + X_t^{ns}, \qquad (23a)$$
$$Y_t = Y_t^{nd} + Y_t^{ns}. \qquad (23b)$$

By statements (22), it can be seen that $X_{t_1}^{nd}$ and $X_{t_2}^{ns}$ are uncorrelated for any $t_1, t_2 \in \mathbb{Z}_+$. In other words, the non-stationary stochastic subsystem will be uncorrelated with its noise-free complement. Furthermore, this non-stationary stochastic subsystem can be envisaged as the original stochastic subsystem except that the initial condition is changed from X_0^s to $X_0^{ns} := X_0^s + \tilde{X}_0$.

In addition, with statement (22a) it is evident that $\tilde{X}_0$ and $\{X_t^s, Y_t^s\}$ are uncorrelated. We introduce a new matrix $\tilde{G}_t$, for $t \in \mathbb{N}$, as follows

$$\tilde{G}_t := \frac{1}{q}\mathbf{E}([X_{t+1}^{ns}][Y_t^{ns}]^*)$$

$$= \frac{1}{q}\mathbf{E}([X_{t+1}^s + A^{t+1}\tilde{X}_0][Y_t^s + CA^t\tilde{X}_0]^*).$$

Again, q is the length of each individual signal sequence. Now, define $\tilde{P}_0 := (1/q)\mathbf{E}[\tilde{X}_0\tilde{X}_0^*]$. Then,

$$\tilde{G}_t = A\Sigma^s C^* + \Sigma^{wv} + A^{t+1}\tilde{P}_0(A^t)^*C^*$$

$$= A[\Sigma^s + A^t\tilde{P}_0(A^t)^*]C^* + \Sigma^{wv}$$

$$= A\tilde{\Sigma}_t^s C^* + \Sigma^{wv}, \qquad (24)$$

where $\tilde{\Sigma}_t^s = \Sigma^s + A^t\tilde{P}_0(A^t)^*$ is the non-stationary stochastic state covariance matrix. Note that $\tilde{G}_t$ plays a similar role as G, which, with A, forms the non-stationary stochastic controllability pair $(A, \tilde{G}_t)$.

6.2. *Decomposition with a Kalman filter subsystem*

Again, decompose X_0 into $\hat{X}_0 + \tilde{X}_0 + X_0^s$, as in equations (21). Define $\hat{X}_t$ and $\hat{Y}_t$ recursively from the following Kalman filter subsystem

$$\hat{X}_{t+1} = (A - K_t C)\hat{X}_t + (B - K_t D)U_t + K_t Y_t,$$
$$\hat{Y}_t = C\hat{X}_t + DU_t,$$

where K_t and $\tilde{P}_0$ are defined in equations (11). Now denote by $\tilde{X}_t^s$ the error of the Kalman filter state, and by $\tilde{Y}_t^s$ the error of the Kalman filter output

$$\tilde{X}_t^s := X_t - \hat{X}_t := \tilde{X}_t + X_t^s,$$
$$\tilde{Y}_t^s := Y_t - \hat{Y}_t.$$

It can be shown that $\tilde{X}_t^s$ and $\tilde{Y}_t^s$ satisfy the following complementary error subsystem

$$\tilde{X}_{t+1}^s = (A - K_t C)\tilde{X}_t^s - K_t V_t + W_t;$$
$$\tilde{Y}_t^s = C\tilde{X}_t^s + V_t.$$

For convenience, adopt, for $t_2 > t_1$, the notation

$$\Phi(t_2, t_1) := (A - K_{t_2-1}C)(A - K_{t_2-2}C)\cdots(A - K_{t_1}C). \tag{25}$$

7. Spaces of input and projected state sequences

This section provides the generalisation of §4 that is required to treat the combined case. Due to the presence of stochastic disturbances, system equations are projected onto some instrumental variable space $\mathcal{S}$, so as to eliminate the effect of stochastic noises. It has been seen that this instrumental variable space is usually taken to be $\mathcal{Y}_f + \mathcal{U}_p + \mathcal{U}_f$ in subspace methods. Owing to this fact, the emptiness of the intersection of $\Pi_{\mathcal{S}} X_f$ and U_f is desired. More precise statements of this condition are given in the following proposition in various forms.

Proposition 2: *For $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$, the following are equivalent*

1. $\mathbf{E}[\mathcal{U}_f \mathcal{U}_f^*] > 0$, and $\dim\left\{\Pi_{\mathcal{S}}\Pi_{\mathcal{U}_f^\perp} X_f\right\} = n$.
2. $\mathbf{E}[\mathcal{U}_f \mathcal{U}_f^*] > 0$, and $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}} X_f\right\} = n$.
3. $\mathbf{E}\left[\begin{pmatrix}\Pi_{\mathcal{S}} X_f \\ \mathcal{U}_f\end{pmatrix}\begin{pmatrix}\Pi_{\mathcal{S}} X_f \\ \mathcal{U}_f\end{pmatrix}^*\right] > 0$.
4. $\mathbf{E}\left[\begin{pmatrix}X_f \\ \mathcal{U}_f\end{pmatrix}(Y_p^* \quad U_p^* \quad U_f^*)\right]$ *is of full row rank.*

The proof of this proposition can be found in Appendix A. This section is organised into three subsections as follows. Section 7.1 covers the case where

$\Sigma^w = 0$, while §7.2 covers the case where $(A, [\Sigma^w]^{1/2})$ is reachable. Finally, §7.3 presents the generalized case where $(A, [B \ (\Sigma^w)^{1/2}])$ is reachable. It is emphasised that if an 'averaging' operator $\bar{\mathbf{E}}$ is used instead of an expectation operator $\mathbf{E}$, only consistent results can be obtained.

7.1. *Stochastic output disturbance*

Systems with only stochastic output disturbance are also known as output error systems. For such systems, we simply have $\Sigma^w = 0$. Since internal noise excitation is absent from the system, excitation relies on the external input signal; in which case, (A, B) being reachable is required. First, a straightforward lemma is presented.

Lemma 1: *Let (A, B) be reachable. Then, $\begin{bmatrix}A^k & A^{k-1}B & \cdots & B\end{bmatrix}$ is of full row rank, for any $k \geq 1$.*

The proof of this lemma can be found in Appendix A. Note that this lemma, together with material derived in §4, are the key point tightening the input excitation requirement from $k + l + n$ to $k + l + r$. The next theorem gives an input excitation requirement such that the conditions in Proposition 2 are satisfied. Note that these requirements are exactly the same as those in Theorem 5 for the noise-free case. In other words, input excitation is again the key to obtaining informative experiments.

Theorem 6: *Let Z^N be an input–output signal pair of a system* (1), *with*

1. $\Sigma^w = 0$ *and (A, B) reachable;*
2. *r no smaller than the degree of the A-annihilator;*
3. *k no smaller than the observability index of the system;*
4. *the input signal u S.R. of order $k + l + r$.*

Then, $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}} X_f\right\} = n$, *where $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$.*

Proof: Note that $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$ is equal to the space spanned by U_p, U_f and $(Y_p - \mathcal{T}_k^d U_p) = (\mathcal{O}_k X_p + V_p)$. Thus, by Proposition 2 it is equivalent to show that

$$\mathbf{E}\left[\begin{pmatrix}X_f \\ \mathcal{U}_f\end{pmatrix}\left([\mathcal{O}_k X_p + V_p]^*, U_p^*, U_f^*\right)\right] = \mathbf{E}\left[\begin{pmatrix}X_f \\ U_f\end{pmatrix}(X_p^*\mathcal{O}_k^*, U_p^*, U_f^*)\right]$$

is of full row rank. Note that the above equality stems from the fact that V_p is orthogonal to X_f and U_f. Finally, realising that $X_f = A^k X_p + \mathcal{C}_k^d U_p$ gives

$$\mathbf{E}\left[\begin{pmatrix}X_f \\ U_f\end{pmatrix}\left(X_p^*\mathcal{O}_k^* \ U_p^* \ U_f^*\right)\right]$$

$$= \begin{bmatrix}A^k & \mathcal{C}_k^d & 0 \\ 0 & 0 & I\end{bmatrix}\mathbf{E}\left[\begin{pmatrix}X_p \\ U_p \\ U_f\end{pmatrix}\begin{pmatrix}X_p \\ U_p \\ U_f\end{pmatrix}^*\right]\begin{bmatrix}\mathcal{O}_k^* & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I\end{bmatrix}.$$

Note that the covariance matrix above is positive definite by Theorem 2. In addition, the first term is of full row rank by Lemma 1. Thus, the above expression, as a product of full row rank matrices, is also of full row rank.   □

7.2. *Stochastic state excitation*

This subsection covers the case where $(A, [\Sigma^w]^{1/2})$ is reachable. In other words, this corresponds to systems whose states are fully stochastically excited. For such systems, the stochastic controllability pair (A, G) is required to be reachable. Also because of the introduction of this process noise, perfect noise-free/stochastic cancellation may occur. Such a phenomenon is also pointed out in Jansson and Wahlberg (1998), in which a counterexample demonstrating this effect is given. In the following, an exact condition preventing such cancellation is derived.

First of all, it is shown that under the assumption that the stochastic subsystem is stationary, the system matrix A must be asymptotically stable.

Lemma 2 (Pernebo and Silverman 1982)**:** *Suppose* $(A, [\Sigma^w]^{1/2})$ *reachable. Then, A is asymptotically stable if and only if*

$$0 < \Sigma^s = A\Sigma^s A^* + \Sigma^w < \infty.$$

$$\Psi := \begin{pmatrix} A^{k-r-1}G + A^k\tilde{P}_0(A^r)^*C^* & A^{k-r-2}G + A^k\tilde{P}_0(A^{r+1})^*C^* & \cdots & G + A^k\tilde{P}_0(A^{k-1})^*C^* \\ A^{k-r}G + A^k\tilde{P}_0(A^{r-1})^*C^* & A^{k-r-1}G + A^k\tilde{P}_0(A^r)^*C^* & \cdots & AG + A^k\tilde{P}_0(A^{k-2})^*C^* \\ \ddots & \ddots & \ddots & \ddots \\ A^{k-1}G + A^k\tilde{P}_0C^* & A^{k-2}G + A^k\tilde{P}_0A^*C^* & \cdots & A^rG + A^k\tilde{P}_0(A^{k-r-1})^*C^* \end{pmatrix}.$$

Before proving Theorem 7 regarding the controllability of some non-steady stochastic pairs, the following lemma is presented first.

Lemma 3: *Let* $p_{\min}(z) = z^r + a_{r-1}z^{r-1} + \cdots + a_0$ *be the A-annihilator, where A is asymptotically stable. Then, the $n \times n$ matrix $I + a_{r-1}A + \cdots + a_0A^r$ is non-singular.*

Proof: Suppose $p_{\min}(z)$ can be factorised as $p_{\min}(z) = (z - s_1)(z - s_2)\cdots(z - s_r)$, where $s_1, \ldots, s_r$ are eigenvalues of A. Then, we have

$$I + a_{r-1}A + \cdots + a_0A^r = (I - s_1A)(I - s_2A)\cdots(I - s_rA).$$

Therefore, $I + a_{r-1}A + \cdots + a_0A^r$ is singular if one of the $(I - s_jA)$ is singular. Now suppose $s_j = 0$, then certainly $(I - s_jA) = I$ is not singular. On the other hand,

suppose $s_j \neq 0$. Then, $(I - s_jA)$ is singular if and only if $(s_j^{-1}I - A)$ is singular, or equivalently, s_j^{-1} is an eigenvalue of A. However, since A is asymptotically stable, we have $|s_j| < 1$. Therefore, s_j^{-1}, where $|s_j^{-1}| > 1$, cannot be an eigenvalue of A. In other words, each of $(I - s_jA)$ is non-singular; and hence, the result follows.   □

In the following, the controllability of the non-steady pairs $(A, \tilde{G}_t)$ and $(\Phi(t_2, t_1), K_{t_1})$, defined in §6, is investigated. These pairs, changing with respect to time, can be seen as time-varying stochastic controllability pairs. The following theorem shows that if a long enough time period is allowed, these pairs are in fact reachable.

Theorem 7: *Suppose each of the following holds*:

1. $(A, [\Sigma^w]^{1/2})$ *and* (A, G) *are reachable*;
2. *k is no smaller than twice the degree of the A-annihilator.*

Then for $\tilde{G}$ and $\Phi(k, \cdot)$ respectively defined in equations (24) and (25),

1. $[A^{k-1}\tilde{G}_0 \quad A^{k-2}\tilde{G}_1 \quad \cdots \quad \tilde{G}_{k-1}]$ *is of full row rank*;
2. $[\Phi(k, 0)K_0 \quad \Phi(k, 1)K_1 \quad \cdots \quad K_{k-1}]$ *is of full row rank.*

Proof: Let r be the degree of the A-annihilator and $p_{\min}(z) = z^r + a_{r-1}z^{r-1} + \cdots + a_0$ the A-annihilator. So, we have $k \geq 2r$. Define the Toeplitz matrix Ψ as:

Since $p_{\min}(A^*) = 0$, we have

$$A^k\tilde{P}_0 \cdot p_{\min}(A^*) \cdot (A^h)^*C^* = 0,$$

for any $h \in \mathbb{Z}_+$. Thus,

$$[I \quad a_{r-1}I \quad \cdots \quad a_0I]\Psi$$
$$= [I \quad a_{r-1}I \quad \cdots \quad a_0I]$$
$$\times \begin{pmatrix} A^{k-r-1}G & A^{k-r-2}G & \cdots & G \\ A^{k-r}G & A^{k-r-1}G & \cdots & AG \\ \ddots & \ddots & \ddots & \ddots \\ A^{k-1}G & A^{k-2}G & \cdots & A^rG \end{pmatrix}$$
$$= (I + a_{r-1}A + \cdots + a_0A^r)[A^{k-r-1}G \quad A^{k-r-2}G \quad \cdots \quad G].$$

As shown in Lemma 2, A must be asymptotically stable for any stationary stochastic process. Consequently,

we have $(I + a_{r-1}A + \cdots + a_0 A^r)$ non-singular by Lemma 3. Finally, since $k - r - 1 \geq r - 1$ and (A, G) is reachable, we have

$$\text{rank}\left\{\left[A^{k-r-1}G \; A^{k-r-2}G \; \cdots \; G\right]\right\} = n.$$

Now, it is easy to see that

$$\text{rank}\left\{[I \; a_{r-1}I \; \cdots \; a_0 I]\Psi\right\} = n. \tag{26}$$

1. Statement 1 will be proven by contradiction. Suppose there exists $\nu \neq 0$ such that

$$\nu\left[A^{k-1}\tilde{G}_0 \; A^{k-2}\tilde{G}_1 \; \ldots \; \tilde{G}_{k-1}\right] = 0.$$

This implies $\nu\left(A^{k-j}G + A^k\tilde{P}_0(A^{j-1})^*C^*\right) = 0$, for $j = 1, \ldots, k$. Thus,

$$\nu[I \; a_{r-1}I \; \cdots \; a_0 I]\Psi = [\nu \; a_{r-1}\nu \; \cdots \; a_0\nu]\Psi = 0,$$

which is clearly a contradiction to equation 26.

2. By Theorem 1 we write

$$\begin{aligned}
\mathcal{K}_k &= [(A - K_{k-1}C)\mathcal{K}_{k-1} \; K_{k-1}] \\
&= [(A - K_{k-1}C)(A - K_{k-2}C)\mathcal{K}_{k-2} \\
&\quad (A - K_{k-1}C)K_{k-2} \; K_{k-1}] \\
&= [\Phi(k,0)K_0 \; \Phi(k,1)K_1 \; \cdots \; K_{k-1.}]
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\mathcal{K}_k &= (A^k\tilde{P}_0\mathcal{O}_k^* + \mathcal{C}_k^s)(L_k + \mathcal{O}_k\tilde{P}_0\mathcal{O}_k^*)^{-1} \\
&= \begin{bmatrix} A^k(\Sigma^s + \tilde{P}_0)C^* + A^{k-1}\Sigma^{wv} \; A^{k-1}(\Sigma^s + A\tilde{P}_0A^*)C^* \\ + A^{k-2}\Sigma^{wv} \; \cdots \; A(\Sigma^s + A^{k-1}\tilde{P}_0A^{k-1})C^* + \Sigma^{wv} \end{bmatrix} \\
&\quad \times (L_k + \mathcal{O}_k\tilde{P}_0\mathcal{O}_k^*)^{-1} \\
&= \left[A^{k-1}\tilde{G}_0 \; A^{k-2}\tilde{G}_1 \; \cdots \; \tilde{G}_{k-1}\right](L_k + \mathcal{O}_k\tilde{P}_0\mathcal{O}_k^*)^{-1}.
\end{aligned}$$

Note that since $(A, [\Sigma^w]^{1/2})$ is controllable, Σ^s is positive definite. Without loss of generality, this system can be regarded as the reduced stochastic system defined in equation (10). Note also that with strictly positive realness, L_k is guaranteed to be positive definite, as shown in Faurre (1976). Thus, statement 1 and statement 2 are equivalent.

This completes the proof. □

The next theorem provides the conditions under which the statements in Proposition 2 will be satisfied in the case of full stochastic state excitation. Unlike the noise-free case in which a certain degree of input excitation is required, we require in this case a longer past data block to avoid rank cancellation, similar to that shown in Theorem 7. This ensures the emptiness of the intersection of the input and projected state sequences.

Theorem 8: *Let* Z^N *be an input–output signal pair of a system* (1), *with*

1. $(A, (\Sigma^w)^{1/2})$ *and* (A, G) *reachable*;
2. k *no smaller than twice the degree of the A-annihilator*;
3. *the input signal u S.R. of order* $k + l$.

Then, $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_\mathcal{S}X_f\right\} = n$, *where* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$.

The proof of the theorem can be found in Appendix A. From the above theorem, it can be seen that due to the internal state excitation by the process noise, a lower degree of input excitation is required as compared to that in Theorem 5. However, a long enough past data block is required.

7.3. *General excitation and disturbance*

Finally, the development is extended to a much more general input structure, in which no specific constraint is imposed on B and Σ^w as long as these matrices, together with A, form a reachable pair. In other words, such a setup covers cases where Σ^v is not necessarily positive definite and $(A, (\Sigma^w)^{1/2})$ is not necessarily reachable. To facilitate the analysis, transformation as described in Proposition 1 is performed. It is clear that requirements for the lower two-block reduced system have been presented in Theorem 6, and the upper one-block reduced system in Theorem 8. Therefore, the difficulty lies in the hybrid combination of the two reduced systems.

Theorem 9: *Let* Z^N *be an input–output signal pair of a system* (1), *with*

1. A *asymptotically stable*;
2. $(A, [B \; (\Sigma^w)^{1/2}])$ *reachable*;
3. $(A, [\Sigma^w]^{1/2})$ *and* (A, G) *having common reachable modes*;[‡]
4. r *no smaller than the degree of the A-annihilator*;
5. k *no smaller than three times the degree of the A-annihilator*;
6. *the input signal u S.R. of order* $k + l + r$.

Then, $\dim\{\Pi_{\mathcal{U}_f^\perp}\Pi_\mathcal{S}X_f\} = n$, *where* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$.

The proof is tedious and can be found in Appendix A.

This theorem implies that additional past data block rows are required to accommodate the further degree of freedom in the system input structure.

[‡]That is, for any $\lambda \in \mathbb{C}$, $\nu \in \mathbb{C}^n$, we have $\nu[A - \lambda I \; (\Sigma^w)^{1/2}] = 0$ if and only if $\nu[A - \lambda I \; G] = 0$.

8. General requirements for informative experiments

This section is concerned with informative criteria for Algorithm 3.2, which is capable of identifying systems in the combined case. Criteria for the three different conditions discussed in the previous section are presented in the following. As in the setup of §5, we shall confine the domain of data sets to all data sets which are generated by nth order systems described by equations (1) with noises $w(t)$ and $v(t)$, and some initial condition $x(0)$.

First of all, note that in Algorithm 3.2, a combined noise-free/stochastic data equation (8b) is converted to a noise-free-like data equation by projecting it onto the subspace $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$:

$$\mathbf{\Pi}_{\mathcal{S}} Y_f = \mathcal{O}_k \mathbf{\Pi}_{\mathcal{S}} X_f + \mathcal{T}_k^d U_f. \tag{27}$$

Effectively, Algorithm 3.2 is identical to Algorithm 3.1 if $\mathbf{\Pi}_{\mathcal{S}} Y_f$ and $\mathbf{\Pi}_{\mathcal{S}} X_f$ are replaced by Y and X_0 respectively. Thus, as seen in Theorems 4 and 5 for the noise-free case, the remaining key point for experiments to be informative is to find a condition such that

$$\dim\left\{\mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_{\mathcal{S}} X_f\right\} = n. \tag{28}$$

First, the informative criteria for the case where $\Sigma^w = 0$ are presented in the following theorem. Before stating the theorem, it should be pointed out that in a real data setup where the 'averaging' operator $\bar{\mathbf{E}}$ is applied, the true system is only estimated consistently.

Theorem 10: *Let Z^N be an input–output signal pair of an observable system* (1), *with*

1. $\Sigma^w = 0$ *and* (A, B) *reachable*;
2. *r no smaller than the degree of the A-annihilator*;
3. $\min\{k, l-1\}$ *no smaller than the observability index of the system*;
4. *the input signal u S.R. of order* $k + l + r$.

Then, Algorithm 2 consistently identifies the system $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T$ *invertible*$\}$.[‡]

Proof: From equation (27) it is easy to see that $\mathcal{O}_l \, \mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_{\mathcal{S}} X_f = \Gamma_1 \Sigma_1 \Omega_1^*$, where Γ_1, Σ_1 and Ω_1 are defined in equation (19). With the rank requirement in equation (28) guaranteed by Theorem 6, we have

$$\mathcal{O}_l = \Gamma_1 \quad \text{and} \quad \mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_{\mathcal{S}} X_f = \Sigma_1 \Omega_1^*$$

in some particular state coordinates. Together with $\mathbf{E}[U_f \, U_f^*] > 0$, the proof is then completed by following that of Theorem 4.   □

Next, the case where $(A, (\Sigma^w)^{1/2})$ is reachable is considered. In this case, the rank requirement in equation (28) is guaranteed by the conditions in Theorem 8. Thus, this theorem can easily be proven by modifying the proof of Theorem 10; and hence the proof is omitted here.

Theorem 11: *Let Z^N be an input–output signal pair of an observable system* (1), *with*

1. $(A, (\Sigma^w)^{1/2})$ *and* (A, G) *reachable*;
2. *k no smaller than twice the degree of the A-annihilator*;
3. $l - 1$ *no smaller than the observability index of the system*;
4. *the input signal u S.R. of order* $k + l$.

Then, Algorithm 3.2 consistently identifies the system $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T$ *invertible*$\}$.

Finally, the general excitation and disturbance case is presented. As expected, the following theorem can also be derived from the conditions stated in Theorem 9; and hence the proof is again omitted here.

Theorem 12: *Let Z^N be an input–output signal pair of an observable system* (1), *with*

1. *A asymptotically stable*;
2. $(A, [B \quad (\Sigma^w)^{1/2}])$ *reachable*;
3. $(A, [\Sigma^w]^{1/2})$ *and* (A, G) *having common reachable modes*;
4. *r no smaller than the degree of the A-annihilator*;
5. *k no smaller than three times the degree of the A-annihilator*;
6. $l - 1$ *no smaller than the observability index of the system*;
7. *the input signal u S.R. of order* $k + l + r$.

Then, Algorithm 2 consistently identifies the system $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T$ *invertible*$\}$.

Remark 3: At this point, attention should be drawn to the fact that the criteria presented in this paper are only sufficient conditions for experiments to be informative. That is, in some cases subspace methods can still identify a system even though these criteria are not satisfied.

Remark 4: Theorem 12 is more general than the results available in Jansson and Wahlberg (1998). In that paper results are obtained for the case that $\Sigma^w = 0$, or for single-input systems, or under the restriction that u is white noise. We do not impose any of these restrictions here.

[‡]The term 'consistently identifies' refers to the fact that, when the 'averaging' operator $\bar{\mathbf{E}}_T$ is used on real data, the identified system converges almost surely to the true system (up to a state-coordinate transformation) as $T \to \infty$

9. Summary

In this paper, we have shown that a rank condition, or the emptiness of the intersection of certain input and projected state sequences, is the key point for constructing informative experiments for subspace algorithms. The requirements for achieving this rank condition are presented

- in §4 for the the noise-free case, where $\Sigma^w = \Sigma^v = 0$;
- in §7.1 for the output disturbance case, where $\Sigma^w = 0$;
- in §7.2 for the stochastic state excitation case, where $(A, [\Sigma^w]^{1/2})$ is reachable;
- in §7.3 for the general excitation and disturbance case, where $(A, [B \ (\Sigma^w)^{1/2}])$ is reachable.

From this analysis, criteria for informative experiments for the two subspace algorithms given in §3 are derived in §5 for the noise-free case, and in §8 for the other three cases.

Our analysis goes beyond that available in Jansson and Wahlberg (1998), in that we do not place any restriction on either the unobserved 'noise'—stochastic excitation of the system—or the observed input. The results obtained in §8 hold for multivariable systems with general stochastic excitation and arbitrary input signals. The conditions we obtain are still sufficient conditions only, however. If rank cancellation occurs then the identified model will almost certainly be incorrect. However rank cancellation may not occur, even if our conditions do not hold. The degree of conservativeness of our results remains a matter for investigation.

The requirements presented in this paper may be considered mild in the sense that they are easily satisfied in typical identification practice. For instance, structural conditions, like the reachability of (A, B) or (A, G), hold in most cases, except in some extreme rank deficient cases. Another requirement is the order of persistent excitation of the input signal, which has to be slightly larger (by the degree of the A-annihilator) than the total number of Hankel block rows. Unless the input signal has certain repeating patterns, its excitation is normally sufficient. In the general excitation and disturbance case, the truncation index k should be larger than three times the degree of the A-annihilator. Such a requirement is mild since, as indicated in Deistler *et al.* (1995), Peternell *et al.* (1996), one should increase the truncation index k at a rate equal to some function involving $\log N$ in order to minimize the BIC criterion, and hence obtain correct estimation of the order n. Therefore, it is reasonable to assume the rank n condition holds by default in a typical subspace identification operation; otherwise not only are these mild requirements violated but also the input-state signals coincidentally create an exact cancellation effect.

10. Acknowledgement

A. Appendix

A.1. *Proof of Theorem 2*

Proof: To begin with, introduce two additional Hankel matrices of signals u and x. Define $\check{U}$ as:

$$\check{U} := \begin{bmatrix} u(0) & u(1) & \cdots & u(N-h-s) \\ u(1) & u(2) & \cdots & u(N-h-s+1) \\ \cdot & \cdot & \cdot & \cdot \\ u(h+s-1) & u(h+s) & \cdots & u(N-1) \end{bmatrix}.$$

This matrix has more rows but fewer columns than U (defined in (6)). Define $\check{X}$ in an analogous way. Let $p_{\min}(z) = z^r + a_{r-1}z^{r-1} + \cdots + a_0$ be the minimal polynomial of A, $(r \le n)$. Now, define Ψ_0 as:

$$\Psi_0 := \check{X}_s + a_{r-1}\check{X}_{s-1} + \cdots + a_0\check{X}_{s-r}. \tag{29}$$

Note that by recursive expansion the state equation has the form

$$\check{X}_{s-j} = A^{r-j} \check{X}_{s-r} + \sum_{i=1}^{r-j} A^{i-1} B \, \check{U}_{s-i-j},$$

where $s \ge r$ as given in the conditions of the theorem. Applying these equalities (for $j = 0, 1, \ldots, r$) to Ψ_0 gives

$$\begin{aligned} \Psi_0 = &(A^{r-1}B + a_{r-1}A^{r-2}B + \cdots + a_1 B) \, \check{U}_{s-r} \\ &+ \cdots + (AB + a_{r-1}B) \, \check{U}_{s-2} + (B) \, \check{U}_{s-1}. \end{aligned}$$

It is noteworthy that the initial state sequence $\check{X}_{s-r}$ is cancelled out since $p_{\min}(A)\check{X}_{s-r} = 0$, by the property of the A-annihilator. On the other hand, define Ψ_j, for $j = 1, 2, \ldots, h$ as

$$\Psi_j := \check{U}_{s+j-1} + a_{r-1}\check{U}_{s+j-2} + \cdots + a_0\check{U}_{s+j-r-1} \tag{30}$$

and let Ψ be the aggregate matrix $\Psi^* := [\Psi_0^* \ \Psi_1^* \ \cdots \ \Psi_h^*]$. By grouping similar terms, it is not hard to see that

$$\Psi = \Upsilon_1 \Upsilon_2 \, \breve{U}_{0,h+s-1},$$

where Υ_1 and Υ_2 are defined as follows:

$$\Upsilon_1 := \begin{bmatrix} A^{r-1}B & A^{r-2}B & \cdots & B & 0 \\ 0 & 0 & 0 & 0 & I \end{bmatrix},$$

$$\Upsilon_2 := \begin{bmatrix} I & & & & & 0 \\ a_{r-1}I & \ddots & & & & \\ \vdots & \ddots & \ddots & I & & \\ a_0I & & a_{r-1}I & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & a_0I & \cdots & a_{r-1}I & I \end{bmatrix}.$$

The key point in this proof is to observe that Υ_1 is of full row rank since (A, B) is assumed to be controllable. In addition, Υ_2 clearly has full row rank. Moreover, note that by the definition of sufficient richness, $\mathbf{E}[\breve{U}\breve{U}^*]$ and all its principal submatrices are positive definite. Thus, taking the correlation of Ψ yields a positive definite matrix:

$$\mathbf{E}\left[\Psi\Psi^*\right] = \Upsilon_1 \Upsilon_2 \, \mathbf{E}\left[\breve{U}_{s-r,h+s-1}\breve{U}_{s-r,h+s-1}^*\right] \Upsilon_2^* \Upsilon_1^* > 0. \tag{31}$$

The theorem statement will be proven by contradiction. Suppose the rows of $\mathbf{\Pi}_{\mathcal{U}^\perp} X_0$ span a subspace with dimension less than n. There exist nonzero constant $\alpha_0 \in \mathbb{R}^n$ and constants $\alpha_1, \ldots, \alpha_h \in \mathbb{R}^m$ such that

$$\alpha_0^* X_0 + \alpha_1^* U_0 + \cdots + \alpha_h^* U_{h-1} = 0.$$

Expanding each component of X_0 and U_i gives

$$\alpha_0^* x(t) + \alpha_1^* u(t) + \cdots + \alpha_h^* u(t+h-1) = 0,$$

for $t = 0, 1, \ldots, N-h$. Grouping these equalities together gives

$$\Theta_j := \alpha_0^* \breve{X}_j + \alpha_1^* \hat{U}_j + \cdots + \alpha_h^* \hat{U}_{h+j-1} = 0, \tag{32}$$

for $j = 0, 1, \ldots, s$. Note that each of these identities is denoted by Θ_j. In addition, define $\alpha^* := [\alpha_0^* \ \alpha_1^* \ \cdots \ \alpha_h^*]$. From equation (32) together with

equations (29) and (30), the following equality can then be derived:

$$\begin{aligned} \alpha^* \Psi &= \alpha_0^* \Psi_0 + \alpha_1^* \Psi_1 + \cdots + \alpha_h^* \Psi_h \\ &= \Theta_s + a_{r-1}\Theta_{s-1} + \cdots + a_0 \Theta_{s-r} \\ &= 0. \end{aligned}$$

However, taking the correlation gives $\alpha^* \mathbf{E}[\Psi\Psi^*]\alpha = 0$, which clearly contradicts equation (31). Therefore, $\mathbf{\Pi}_{\mathcal{U}^\perp} \mathcal{X}_0$ is of dimension n. This completes the proof of the theorem. $\qquad\square$

A.2. *Proof of Theorem 3*

Proof: First, let e be a noise process such that e is uncorrelated with u, and

$$\mathbf{E}(e(t)e^*(\tau)) = I\delta_{t\tau}, \qquad w(t) = (\Sigma^w)^{1/2}e(t).$$

Define

$$\breve{B} = \begin{bmatrix} B & (\Sigma^w)^{1/2} \end{bmatrix}, \quad \text{and} \quad \breve{u}(t) := [u(t)^* \ e(t)^*]^*.$$

Now, the theorem can be proven using Theorem 2 by replacing B by $\breve{B}$ and $u(t)$ by $\breve{u}(t)$. $\qquad\square$

A.3. *Proof of Theorem 4*

Proof: The proof will be based on Theorem 2, which guarantees $\mathbf{\Pi}_{\mathcal{U}^\perp} \mathcal{Y}$ is of dimension n. Consider equation (8b). Projecting each row onto the space $\mathcal{U}^\perp$ yields

$$\mathbf{\Pi}_{\mathcal{U}^\perp} Y = \mathcal{O}_h \mathbf{\Pi}_{\mathcal{U}^\perp} X_0,$$

which cancels the effect of the input Hankel block U.

Since $\mathcal{O}_h$ is injective, the rows of $\mathbf{\Pi}_{\mathcal{U}^\perp} Y$ span an n-dimensional subspace, by Theorem 2. Suppose we choose certain state coordinates (note that we allow a similarity transformation $GL_n(\mathbb{R})$) such that

$$\mathbf{E}[(\mathbf{\Pi}_{\mathcal{U}^\perp} X_0)(\mathbf{\Pi}_{\mathcal{U}^\perp} X_0)^*] = \Sigma_1 \Sigma_1^*,$$

with Σ_1 as defined in equation (15). Then we can take

$$\mathcal{O}_h = \Gamma_1 \quad \text{and} \quad \mathbf{\Pi}_{\mathcal{U}^\perp} X_0 = \Sigma_1 \Omega_1^*,$$

with Γ_1 and Ω_1^* as defined in equation (15). Thus, with the injectivity of $\mathcal{O}_{h-1}$, A is uniquely determined by the shift invariant property and C is uniquely determined from the first p rows of $\mathcal{O}_h$, as shown in Step 3 of Algorithm 3.1.

Note that since $\mathcal{O}_{h-1}$ is injective, there always exists P such that

$$P\mathcal{O}_{h-1} + CA^{h-1} = 0.$$

Define $\mathcal{O}_h^\perp := [P \quad I]$. It is obvious that $\mathcal{O}_h^\perp \mathcal{O}_h = 0$. Furthermore, by the definition of sufficient richness we have $U_0, \ldots, U_{h-1}$ mutually linearly independent. Thus, $S_0, \ldots, S_{h-1}$ in equation (16) are uniquely determined. Partition $\mathcal{O}_h^\perp$ into

$$\mathcal{O}_h^\perp = \begin{bmatrix} P_0 & \cdots & P_{h-2} & I \end{bmatrix}.$$

It is then easy to see that equation (17) holds, which uniquely determines B and D. This completes the proof. $\quad\square$

A.4. *Proof of Proposition 2*

Proof:

(1) $\Leftrightarrow$ (2): We shall prove $\mathbf{\Pi}_{\mathcal{U}_f^\perp}$ and $\mathbf{\Pi}_\mathcal{S}$ commute. Since $\mathcal{U}_f \subset \mathcal{S}$, we have $\mathcal{S}^\perp \subset \mathcal{U}_f^\perp$. This gives

$$\mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_{\mathcal{S}^\perp} = \mathbf{\Pi}_{\mathcal{S}^\perp} = \mathbf{\Pi}_{\mathcal{S}^\perp} \mathbf{\Pi}_{\mathcal{U}_f^\perp}. \tag{33}$$

Since $\mathbf{\Pi}_\mathcal{S} + \mathbf{\Pi}_{\mathcal{S}^\perp} = I$, it implies $(\mathbf{\Pi}_\mathcal{S} + \mathbf{\Pi}_{\mathcal{S}^\perp})\mathbf{\Pi}_{\mathcal{U}_f^\perp} = \mathbf{\Pi}_{\mathcal{U}_f^\perp}(\mathbf{\Pi}_\mathcal{S} + \mathbf{\Pi}_{\mathcal{S}^\perp})$. Using equation (33), we have $\mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_\mathcal{S} = \mathbf{\Pi}_\mathcal{S} \mathbf{\Pi}_{\mathcal{U}_f^\perp}$.

(2) $\Rightarrow$ (3): We shall prove by contradiction. Suppose statement 3 is not true, then there exist column vectors $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^{km}$, not both zero, such that $\alpha^* \mathbf{\Pi}_\mathcal{S} X_f + \beta^* U_f = 0$. This implies the $\mathcal{U}_f^\perp$ component must satisfy:

$$\alpha^* \mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_\mathcal{S} \mathcal{X}_f = 0.$$

Thus, if $\alpha \neq 0$, this violates $\dim\{\mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathbf{\Pi}_\mathcal{S} X_f\} = n$. On the other hand, suppose $\alpha = 0$. Then, $\beta \neq 0$ and $\beta^* U_f = 0$ violate $\mathbf{E}[U_f U_f^*] > 0$.

(3) $\Rightarrow$ (4): Note that since $\mathbf{E}[(\mathbf{\Pi}_{\mathcal{S}^\perp} X_f)(Y_p^* \quad U_p^* \quad U_f^*)] = \mathbf{E}[(\mathbf{\Pi}_{\mathcal{S}^\perp} U_f)(Y_p^* \quad U_p^* \quad U_f^*)] = 0$, we have

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} (Y_p^* \quad U_p^* \quad U_f^*) \right]$$
$$= \mathbf{E}\left[\begin{pmatrix} \mathbf{\Pi}_\mathcal{S} X_f \\ \mathbf{\Pi}_\mathcal{S} U_f \end{pmatrix} (Y_p^* \quad U_p^* \quad U_f^*) \right]. \tag{34}$$

Now, we shall prove by contradiction. Recall that $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$. Thus, equation (34) is not of full row rank if and only if there exist $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^{km}$, not both zero, such that $\alpha^* \mathbf{\Pi}_\mathcal{S} X_f + \beta^* U_f = 0$. This implies that statement 3 is not valid.

(4) $\Rightarrow$ (1): Let $S^* = [Y_p^* \quad U_p^* \quad U_f^*]$. Note that

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} (Y_p^* \quad U_P^* \quad U_f^*) \right]$$
$$= \mathbf{E}\left[\begin{pmatrix} \mathbf{\Pi}_{\mathcal{U}_f} X_f \\ U_f \end{pmatrix} S^* \right] + \mathbf{E}\left[\begin{pmatrix} \mathbf{\Pi}_{\mathcal{U}_f^\perp} X_f \\ 0 \end{pmatrix} S^* \right].$$

Since the above equation is of full row rank, the first term implies $\mathbf{E}[U_f S^*]$ is of full row rank, or equivalently, $\mathbf{E}[U_f U_f^*] > 0$. On the other hand, since $\mathbf{\Pi}_{\mathcal{U}_f} \mathcal{X}_f \subset \mathcal{U}_f$, the second term in turn implies $\mathbf{E}[(\mathbf{\Pi}_{\mathcal{U}_f^\perp} X_f) S^*]$ is of full row rank. This is also to say that $\mathbf{E}[(\mathbf{\Pi}_\mathcal{S} \mathbf{\Pi}_{\mathcal{U}_f^\perp} X_f) S^*]$ is of full row rank, which guarantees $\mathbf{\Pi}_\mathcal{S} \mathbf{\Pi}_{\mathcal{U}_f^\perp} \mathcal{X}_f$ to be of dimension n.

This completes the proof. $\quad\square$

A.5. *Proof of Lemma 1*

Proof: Let $\nu \neq 0$ such that

$$\nu \begin{bmatrix} A^k & A^{k-1}B & \cdots & B \end{bmatrix} = 0.$$

Since $\nu A^k = 0$, we have $\nu A^{k+i}B = 0$ for all $i \in \mathbb{N}$. Thus,

$$\nu \begin{bmatrix} A^i B & A^{i-1}B & \cdots & B \end{bmatrix} = 0,$$

for all $i \in \mathbb{N}$, which is a contradiction. $\quad\square$

A.6. *Proof of Theorem 8*

Proof: Let $\hat{X}_p := \mathbf{\Pi}_{\mathcal{U}_p + \mathcal{U}_f} X_p$ and $\tilde{X}_p := X_p^d - \hat{X}_p$. Thus,

$$\mathbf{\Pi}_{(\mathcal{U}_p + \mathcal{U}_f)^\perp} X_p = \tilde{X}_p + X_p^s.$$

It is easy to verify that

$$\mathbf{\Pi}_{(\mathcal{U}_p + \mathcal{U}_f)^\perp} Y_p = \mathbf{\Pi}_{(\mathcal{U}_p + \mathcal{U}_f)^\perp}(\mathcal{O}_k X_p^d + \mathcal{C}_k^d U_p + Y_p^s)$$
$$= \mathcal{O}_k \tilde{X}_p + \mathrm{Y}_p^s.$$

Observe that $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f = \mathbf{\Pi}_{(\mathcal{U}_p + \mathcal{U}_f)^\perp} \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$. Thus, by Lemma 1, we shall prove

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} \left(\begin{bmatrix} \mathcal{O}_k \tilde{X}_p + Y_p^s \end{bmatrix}^* U_p^* \quad U_f^* \right) \right]$$

is of full row rank. It is obvious that $\mathbf{E}[(U_f)(U_p^* \quad U_f^*)]$ is of full row rank by the definition of sufficient richness. Thus, it is sufficient to show $\mathbf{E}[(X_f)(\mathcal{O}_k \tilde{X}_p + Y_p^s)^*]$ is of full row rank. Since $\hat{X}_p$ and U_p are orthogonal to

$\tilde{X}_p$ and Y_p^s, we can write

$$\frac{1}{q}\mathbf{E}[(X_f)(\mathcal{O}_k\tilde{X}_p + Y_p^s)^*] = \frac{1}{q}\mathbf{E}[(A^k X_p^d + \mathcal{C}_k^d U_p + X_f^s) \\ \times (\mathcal{O}_k\tilde{X}_p + Y_p^s)^*] \\ = \frac{1}{q}\mathbf{E}[(A^k\tilde{X}_p + X_f^s)(\mathcal{O}_k\tilde{X}_p + Y_p^s)^*] \\ = A^k\tilde{P}_0\mathcal{O}_k^* + \mathcal{C}_k^s,$$

where $\tilde{P}_0 = (1/q)\mathbf{E}[\tilde{X}_p\tilde{X}_p^*]$ and $q := N-k-l+1$ is the number of filters in the Kalman filter bank. However, we also have

$$A^k\tilde{P}_0\mathcal{O}_k^* + \mathcal{C}_k^s = [A^k(\Sigma^s + \tilde{P}_0)C^* + A^{k-1}\Sigma^{wv} \quad \cdots \\ A(\Sigma^s + A^{k-1}\tilde{P}_0 A^{k-1})C^* + \Sigma^{wv}] \\ = [A^{k-1}\tilde{G}_0 \quad A^{k-2}\tilde{G}_1 \quad \cdots \quad \tilde{G}_{k-1}].$$

Finally, the above is of full row rank by Theorem 7. $\square$

A.7. *Proof of Theorem 9*

Proof: Without loss of generality, we may assume the system has a realization satisfying equations (9), such that $\bar{X}_{t;1}$ corresponds to the stochastically excited modes, $\bar{X}_{t;2}$ to the output disturbance modes, and $\bar{X}_{t;3}$ to the purely deterministic modes. Now, partition X_p and X_f as:

$$X_p = \begin{bmatrix} X_{p;1} \\ X_{p;2} \\ X_{p;3} \end{bmatrix} = \begin{bmatrix} \bar{X}_{p;1} \\ \bar{X}_{p;2} \\ \bar{X}_{p;3} \end{bmatrix} \quad \text{and} \quad X_f = \begin{bmatrix} X_{f;1} \\ X_{f;2} \\ X_f;3 \end{bmatrix} = \begin{bmatrix} \bar{X}_{f;1} \\ \bar{X}_{f;2} \\ \bar{X}_{f;3} \end{bmatrix},$$

Furthermore, we may assume without loss of generality that $k = 3j$, where j is no smaller than the degree of the A-annihilator. Noticing that $Y_p^* = [Y_{0,j-1}^* \; Y_{j,3j-1}^*]$, it is equivalent to prove as indicated by Lemma 1 that

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}(Y_p^* \quad U_p^* \quad U_f^*)\right] \\ = \mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}(Y_{0,j-1}^* \quad Y_{j,3j-1}^* \quad U_p^* \quad U_f^*)\right] \quad (35)$$

is of full row rank. Now let $\hat{X}_j := \Pi_{(\mathcal{U}_p+\mathcal{U}_f)}X_j$ and $\tilde{X}_j := X_j^d - \hat{X}_j$ such that

$$\Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp}X_j = \tilde{X}_j + X_j^s \quad \text{and}$$
$$\Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp}Y_{j,3j-1} = \mathcal{O}_{2j}\tilde{X}_j + Y_{j,3j-1}^s.$$

Note further that $\mathcal{S}$ can be written as

$$\mathcal{S} := \mathcal{U}_p + \mathcal{U}_f + \mathcal{Y}_{0,j-1} + \mathcal{Y}_{j,3j-1} \\ = \mathcal{U}_p + \mathcal{U}_f + \mathcal{Y}_{0,j-1} + \Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp}\mathcal{Y}_{j,3j-1}.$$

As a result, the rank of equation (35) is equivalent to that of

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}\left(Y_{0,j-1}^* \quad U_p^* \quad U_f^* \quad [\mathcal{O}_{2j}\tilde{X}_j + Y_{j,3j-1}^s]^*\right)\right].$$

Now, we consider

$$\mathbf{E}\left[\begin{pmatrix} X_{f;2} \\ X_{f;3} \\ U_f \end{pmatrix}(Y_{0,j-1}^* \quad U_p^* \quad U_f^*)\right] \quad \text{and}$$

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}(\mathcal{O}_{2j}\tilde{X}_j + Y_{j,3j-1}^s)^*\right]. \quad (36)$$

1. To show the first expression in (36) is of full row rank, we follow the proof of Theorem 6. Note that $Y_{0,j-1} - \mathcal{T}_k^d U_{0,j-1} = \mathcal{O}_k X_p + \mathcal{T}_k^w W_{0,j-1} + V_{0,j-1}$ and that $\mathcal{W}_{0,j-1}$ and $\mathcal{V}_{0,j-1}$ are orthogonal to $\mathcal{X}_{f;2}$ and $\mathcal{X}_{f;3}$. Thus, it is equivalent to show

$$\mathbf{E}\left[\begin{pmatrix} X_{f;2} \\ X_{f;3} \\ U_f \end{pmatrix}(X_p^*\mathcal{O}_k^* \quad U_p^* \quad U_f^*)\right] \\ = \begin{bmatrix} A_{OD}^k & \mathcal{C}_{OD,k}^d & 0 \\ 0 & 0 & I \end{bmatrix} \\ \times \mathbf{E}\left[\begin{pmatrix} X_{p;2} \\ X_{p;3} \\ U_p \\ U_f \end{pmatrix}\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}^*\right]\begin{bmatrix} \mathcal{O}_k^* & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \quad (37)$$

is of full row rank, where OD stands for 'Output Disturbance' and

$$A_{OD} = \begin{bmatrix} \bar{A}_{22} & \bar{A}_{23} \\ 0 & \bar{A}_{33} \end{bmatrix}, \quad B_{OD} = \begin{bmatrix} \bar{B}_2 \\ \bar{B}_3 \end{bmatrix},$$
$$[\mathcal{C}_{OD,k}^d = [\bar{A}_{OD}^{k-1}\bar{B}_{OD} \quad \bar{A}_{OD}^{k-2}\bar{B}_{OD} \quad \cdots \quad \bar{B}_{OD}],$$

where B has an obvious partition $B^* = [\bar{B}_1^* \; \bar{B}_2^* \; \bar{B}_3^*]$. Observe by condition 2 that (A_{OD}, B_{OD}) is reachable;

and hence by Theorem 3 we have

$$\mathbf{E}\left[\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}^*\right] > 0,$$

which implies that its principal submatrix is also positive definite

$$\mathbf{E}\left[\begin{pmatrix} X_{p;2} \\ X_{p;3} \\ U_p \\ U_f \end{pmatrix}\begin{pmatrix} X_{p;2} \\ X_{p;3} \\ U_p \\ U_f \end{pmatrix}^*\right] > 0,$$

Now, it can readily be seen that equation 37 is of full row rank since the expression is a product of full row rank matrices.

2. Next, we consider the second expression in (36). Following the proof of Theorem 8, we get

$$\frac{1}{q}\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right]$$
$$= \begin{bmatrix} A^{2j-1}\tilde{G}_0 & A^{2j-2}\tilde{G}_1 & \cdots & \tilde{G}_{2j-1} \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (38)$$

where $q := N-k-l+1$ is the number of filters in the Kalman filter bank. Now, let $p_{\min}(z) = z^j + a_{j-1}z^{j-1} + \cdots + a_0$ be an annihilating polynomial and

$$\Upsilon := \begin{bmatrix} a_0 I & & & 0 \\ \vdots & \ddots & & \\ a_{j-1}I & & a_0 I \\ I & & \ddots & \vdots \\ & & \ddots & a_{j-1}I \\ 0 & & & I \end{bmatrix}.$$

Following the proof of Theorem 7, we obtain

$$\begin{bmatrix} A^{2j-1}\tilde{G}_0 & A^{2j-2}\tilde{G}_1 & \cdots & \tilde{G}_{2j-1} \end{bmatrix}\Upsilon$$
$$= (I + a_{j-1}A + \cdots + a_0 A^j)\begin{bmatrix} A^{j-1}G & A^{j-2}G & \cdots & G \end{bmatrix}. \quad (39)$$

Furthermore, by Lemma 3 we have $(I + a_{j-1}A + \cdots + a_0 A^j)$ non-singular. Thus, together with condition 3 we have

$$\mathbf{E}\left[\left(X_{f;1}\right)\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right]\Upsilon$$

of full row rank. On the other hand, note that

$$A = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix},$$

$$G = A\Sigma^s C^* + \Sigma^{wv} = \begin{bmatrix} \bar{A}_{11}\bar{\Sigma}^s_{11}\bar{C}^*_{11} + \bar{\Sigma}^{wv}_{11} & \bar{\Sigma}^{wv}_{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Using the above equations, it is easy to derive from equation (38) and (39) that

$$\mathbf{E}\left[\begin{pmatrix} X_{f;2} \\ X_{f;3} \\ U_f \end{pmatrix}\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right]\Upsilon = 0.$$

Finally, we complete the proof by observing that

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & 0 \end{bmatrix}$$
$$:= \mathbf{E}\left[\begin{pmatrix} X_{f;1} \\ X_{f;2} \\ X_{f;3} \\ U_f \end{pmatrix}\left(Y^*_{0,j-1} U^*_p U^*_f \left[\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right]^*\right)\right]\begin{bmatrix} I & 0 \\ 0 & \Upsilon \end{bmatrix}$$

is of full row rank since

$$\Delta_{21} := \mathbf{E}\left[\begin{pmatrix} X_{f;2} \\ X_{f;3} \\ U_f \end{pmatrix}\left(Y^*_{0,j-1} \quad U^*_p \quad U^*_f\right)\right]$$

is of full row rank from part (1) and

$$\Delta_{12} := \mathbf{E}\left[\left(X_{f;1}\right)\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right]\Upsilon$$

is of full row rank from part (2). □

References

E. Bai and S. Sastry, "Persistency of excitation, suffcient richness and parameter convergence in discrete time adaptive control", *Systems and Control Letters*, 6, pp. 153–163, 1985.

C. Chou and M. Verhaegen, "Subspace algorithms for the identification of multivariable dynamic errors-in-variables models", *Automatica*, 33, pp. 1857–1869, 1997.

N. Chui, "Subspace Methods and Informative Experiments for System Identification", PhD thesis, Cambridge University, U.K. Also available at URL: http://wwwcontrol.eng.cam.ac.uk/nlcc/report/thesis.ps, 1997.

N. Chui and J. Maciejowski, "Uniqueness of minimal partial realizations and Markov parameter identification", in *35th IEEE Conference on Decision and Control*, Kobe, 1996, pp. 3642–3647.

N. Chui and J. Maciejowski, "Criteria for informative experiments with subspace identification", Technical Report CUED/F-INFENG/TR.338, Cambridge University, U.K. Also available at URL: http://www-control.eng.cam.ac.uk/jmm/papers/TR338.ps, 1998a.

N. Chui and J. Maciejowski, "A geometric interpretation for subspace identification", Technical Report CUED/F-INFENG/TR.339, Cambridge University, U.K. Also available at URL: http://wwwcontrol.eng.cam.ac.uk/jmm/papers/TR339.ps, 1998b.

N. Chui and J. Maciejowski, "Subspace identification algorithm – a Markov parameter approach. Technical Report CUED/F-INFENG/TR.337, Cambridge University, U.K. Also available at URL: http://www-control.eng.cam.ac.uk/jmm/papers/TR337.ps, 1998c.

B. De Moor, Mathematical Concepts and Techniques for Modeling of Static and Dynamic Systems, PhD thesis, Katholieke Universiteit Leuven, Belgium, 1988.

M. Deistler, K. Peternell and W. Scherrer, "Consistency and relative efficiency of subspace methods", *Automatica*, 31, pp. 1865–1875, 1995.

P. Faurre, "Stochastic realization algorithms", in *System Identification: Advances and Case Studies*, R. Mehra and D. Lainiotis, Eds, New York: Academic Press, 1976.

M. Jansson and B. Wahlberg, "On weighting in statespace subspace system identification", in *3rd European Control Conference*, Rome, Italy,1995, Vol. 1, pp. 435–440.

M. Jansson and B. Wahlberg, "On consistency of subspace methods for system identification", *Automatica*, 34(12), pp. 1507–1519, 1998.

T. Kailath, *Linear Systems*, NJ: Prentice-Hall, 1980.

C. Kim and J. Cain, "An *a priori* identifiability condition and order determination algorithm for MIMO systems", *IEEE Transactions on Systems, Man and Cybernetics*, 13, pp. 1119–1128, 1983.

W. Larimore, "Canonical variate analysis in identification, filtering, and adaptive control", in *29th IEEE Conference on Decision and Control*, Honolulu, Hawaii, 1990, pp. 596–604.

L. Ljung, *System Identification – Theory for the Users*, NJ: Prentice-Hall, 1987.

I. Mareels, "Suficiency of excitation", *Systems and Control Letters*, 5, pp. 159–163, 1984.

T. McKelvey and H. Akcay, "Subspace based system identification with periodic excitation signals", *Systems and Control Letters*, 26, pp. 349–361, 1995.

M. Moonen, B. De Moor, L. Vandenberghe and J. Vandewalle, "On- and off-line identification of linear state-space models", *International Journal of Control*, 49, pp. 219–232, 1989.

J. Moore, "Persistence of excitation in extended least squares", *IEEE Transactions on Automatic Control*, 28, pp. 60–68, 1983.

L. Pernebo and L. Silverman, "Model reduction via balanced state space reduction", *IEEE Transactions on Automatic Control*, 27, pp. 382–387, 1982.

K. Peternell, W. Scherrer and M. Deistler, "Statistical analysis of novel subspace identification methods", *Signal Processing*, 52, pp. 161–177, 1996.

P. Stoica and T. Söderström, *System Identification*, NJ: Prentice-Hall, 1989.

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

P. Van Overschee and B. De Moor, "A unifying theorem for three subspace system-identification algorithms", *Automatica*, 31, pp. 1853–1864, 1995.

P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*, The Netherlands: Kluwer Academic Publishers, 1996.

M. Verhaegen, "Subspace model identification–Part 3," *International Journal of Control*, 58, pp. 555–586, 1993.

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

M. Verhaegen and P. Dewilde, "Subspace model identification–Part 1", *International Journal of Control*, 56, pp. 1187–1210, 1992a.

M. Verhaegen and P. Dewilde, "Subspace model identification–Part 2", *International Journal of Control*, 56, pp. 1211–1241, 1992b.

M. Viberg, "Subspace-based methods for the identification of linear time-invariant systems", *Automatica*, 31, pp. 1835–1851, 1995.

B. Wahlberg and M. Jansson, "4SID linear regression", in *33rd IEEE Conference on Decision and Control*, Florida, 1994, pp. 2858–2863.

D. Westwick and M. Verhaegen, "Identifying MIMO Wiener systems using subspace model identification methods", *Signal Processing*, 52, pp. 253–258, 1996.