Consistency analysis of subspace identification methods based on a linear regression approach

Knudsen, Torben

Published in:
Automatica

Publication date:
2001

Document Version
Også kaldet Forlagets PDF

PERGAMON

Automatica 37 (2001) 81–89

automatica

www.elsevier.com/locate/automatica

Brief Paper

Consistency analysis of subspace identification methods based on a linear regression approach☆

Torben Knudsen*

Aalborg University, Institute of Electronic Systems, Department of Control Engineering, Fredrik Bajers Vej 7C, DK-9220 Aalborg Ø, Denmark

Received 22 March 1999; revised 10 April 2000; received in final form 24 May 2000

Abstract

In the literature results can be found which claim consistency for the subspace method under certain quite weak assumptions. Unfortunately, a new result gives a counter example showing inconsistency under these assumptions and then gives new more strict sufficient assumptions which however does not include important model structures such as, e.g. Box–Jenkins. Based on a simple least-squares approach this paper shows the possible inconsistency under the weak assumptions and develops only slightly stricter assumptions sufficient for consistency and which includes any model structure. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Parameter estimation; Linear multivariable systems; Subspace methods; State space models; Consistency; Least-squares methods

1. Introduction

Subspace identification roughly consists of a basic step estimating the extended observability matrix Γ_h and a second step estimating the model parameters.

For the basic step three approaches have recently been in focus. The first relies heavily on linear algebra tools such as orthogonal and oblique projections (van Overschee & De Moor, 1996; Verhaegen, 1993, 1994). The second approach uses the instrumental variable framework (Gustafsson, 1997; Ottersten & Viberg, 1994). In this paper the third least-square (LS) approach (Peternell, Scherrer & Deistler, 1996; Jansson & Wahlberg, 1996) is used as it is more straight forward and gives new insight.

The second parameter estimation step can be based on an estimated state sequence (van Overschee & De Moor, 1996) but the use of states can also be avoided (Verhaegen, 1993, 1994). As the state sequence is merely a result of the measurable input and output it seems natural to derive the estimates without using the intermediate states. All parameter estimates in this paper are based directly on the input–output measurements.

Consistency results are most easily obtained by assuming the number of past samples in the regressor (row dimension in the Hankel matrices) as well as the number of observations in the regression (column dimension in Hankel matrix) to increase to infinity with the number of samples (Peternell et al., 1996). However, the sum of dimensions is limited by the number of samples and it is therefore desirable to keep the row dimension low to allow high column dimension, i.e. many observations in the regression.

Consistency analysis for general systems and a limited Hankel row dimension also exists (van Overschee & De Moor, 1996; Verhaegen, 1994; Jansson & Wahlberg, 1998) but there is a contradiction in the results.

Focus for a while on consistency for $\hat{\Gamma}_h$ because bias on this parameter will propagate to the other parameter estimates. There is then a gap between (van Overschee & De Moor, 1996; Verhaegen, 1994), who claim consistency for $\hat{\Gamma}_h$ if the system is minimal, the input persistently exciting and the number of past samples in the regressor larger or equal to the system order and a very recent result Jansson and Wahlberg (1998) which provides a counterexample showing bias on $\hat{\Gamma}_h$ under the same assumptions. To avoid this bias sufficient conditions are given in Jansson and Wahlberg (1998). These are however a very serious limitation as they do not include Box–Jenkins (BJ)-type models with colored input. Based on the LS approach this paper fills the gap by proving the

☆This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

* Tel.: + 45-96-35-87-46; fax: + 45-98-15-17-39.

E-mail address: tk@control.auc.dk (T. Knudsen).

PII: S0005-1098(00)00125-4

consistency for $\hat{\Gamma}_h$ under these assumptions except that the number of past data in the regressor possibly needs to be larger than the system order but still limited. Under these conditions this paper actually extends the consistency proof from $\hat{\Gamma}_h$ to all parameters in the deterministic part of the model in question in a way which also explains the counterexample given in Jansson and Wahlberg (1998).

The results above do not seem to be available in conference or journal papers. However, one of the anonymous reviewers has kindly supplied the author with a Ph.D. thesis by Chui (1997) which includes similar results based on different approaches and proofs.

The remaining paper is organized as follows. First the problem and notation are introduced. Then the regression model is established and system matrices, e.g. Γ_h are estimated. The consistency analysis for these system matrix estimates are then presented. Next, the model parameters are extracted from the system matrices and consistency results for this step are derived. The results are illustrated with two numerical examples. Finally, the conclusions are drawn.

2. The problem

Subspace identification is used to estimate linear stationary state space models from experimental input and output data. The innovation representation of a state space model is given in Definition 1 and is considered most useful. Below $u_k \in \mathbb{R}^m$ is the input, $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^l$ is the output and $e_k \in \mathbb{R}^l$ is the innovation which are zero mean white noise with covariance R. The order n is assumed known or estimated correctly for which there are methods (Picci, 1997; Sorelius, Söderström, Stoica & Cedervall, 1997; Peternell, 1995; Baur, 1998).

Definition 1 (Innovation model).

$$x_{k+1} = Ax_k + Bu_k + Ke_k,$$

$$y_k = Cx_k + Du_k + e_k,$$

$$\mathrm{E}(e_p e_q^T) \triangleq R\delta_{pq}.$$

The problem is then:

> Given a series of input–output measurements: estimate all the parameters, that is the system matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{l\times n}$ and $D \in \mathbb{R}^{l\times m}$ up to within a similarity transformation and the noise parameters $K \in \mathbb{R}^{n\times l}$, $R \in \mathbb{R}^{l\times l}$ so that the covariance of the output is given by the model.

3. Preliminaries

The basic relation used in the prediction error method (Ljung, 1999) is the recursive state space model relating single samples of the signals. One of the principal new ideas in subspace identification is to combine the recursive state space model into single linear equations relating matrices with parameters to matrices with signals. To do this some definitions are needed.

Definition 2 (*Matrices related to signals*). The input block Hankel matrix is divided into two parts called "past" and "future", where the dimensions are $U_p \in \mathbb{R}^{im\times j}$, $U_f \in \mathbb{R}^{hm\times j}$. Based on the output and innovation there are similar definitions for $Y_p \in \mathbb{R}^{il\times j}$, $Y_f \in \mathbb{R}^{hl\times j}$, $E_p \in \mathbb{R}^{il\times j}$ and $E_f \in \mathbb{R}^{hl\times j}$. The total number of samples used in, e.g. (1) is $N = i + h + j - 1$.

$$\left(\frac{U_p}{U_f}\right) \triangleq \left(\begin{array}{cccc} u_0 & u_1 & \cdots & u_{j-1} \\ u_1 & u_2 & \cdots & u_j \\ \vdots & \vdots & & \vdots \\ u_{i-1} & u_i & \cdots & u_{i+j-2} \\ \hline u_i & u_{i+1} & \cdots & u_{i+j-1} \\ u_{i+1} & u_{i+2} & \cdots & u_{i+j} \\ \vdots & \vdots & & \vdots \\ u_{i+h-1} & u_{i+h} & \cdots & u_{i+h+j-2} \end{array}\right) \tag{1}$$

Remark 2.1. Note that the two parameters i and h allow for a different number of block rows in past U_p and future U_f. This choice is supported in Bauer, Deistler and Scherrer (1997), Gustafsson (1997), Ljung and McKelvey (1997), Ottersten and Viberg (1994), Viberg (1995) while others (Peternell et al., 1996; van Overschee & De Moor, 1996; Deistler, Peternell & Scherrer, 1995; Ohsumi, Takashima & Kameyama, 1997) assume $h = i$ probably for simplicity.

The state matrix X_k is defined as a sequence of states starting from some sample k. Past and future state matrices are defined by $k = 0$ and i, respectively.

$$X_k \triangleq (x_k \quad x_{k+1} \quad \cdots \quad x_{k+j-2} x_{k+j-1}) \in \mathbb{R}^{n\times j},$$

$$X_p \triangleq X_0, \quad X_f \triangleq X_i.$$

A column in a matrix, e.g. Y_f will be denoted with lower letters y_f and $y_f(k)$ if the specific column number is needed. This convention is used for all the signal-related matrices.

Definition 3 (*System matrices related to model parameter*). The extended observability matrix Γ_k is defined as

$$\Gamma_k \triangleq \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{pmatrix} \in \mathbb{R}^{kl\times n}.$$

A generic reversed extended controllability matrices $\mathscr{C}_i$ is defined below where $\mathscr{A}$ and $\mathscr{B}$ represent system and input matrices, respectively.

$$\mathscr{C}_i(\mathscr{A},\mathscr{B}) \triangleq (\mathscr{A}^{i-1}\mathscr{B} \quad \mathscr{A}^{i-2}\mathscr{B} \quad \cdots \quad \mathscr{A}\mathscr{B} \quad \mathscr{B}).$$

Two lower block triangular Toeplitz matrices H_k^{d} and H_k^{s} corresponding to the deterministic and stochastic parts, respectively, are defined below based on the generic block triangular Toeplitz matrices $\mathscr{H}_k$.

$$H_k^{\mathrm{d}} \triangleq \mathscr{H}_k(A,B,C,D),$$

$$H_k^{\mathrm{s}} \triangleq \mathscr{H}_k(A,K,C,I),$$

$$\mathscr{H}_k(\mathscr{A},\mathscr{B},\mathscr{C},\mathscr{D}) \triangleq \begin{pmatrix} \mathscr{D} & 0 & 0 & \cdots & 0 \\ \mathscr{C}\mathscr{B} & \mathscr{D} & 0 & \cdots & 0 \\ \mathscr{C}\mathscr{A}\mathscr{B} & \mathscr{C}\mathscr{B} & \mathscr{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathscr{C}\mathscr{A}^{k-2}\mathscr{B} & \mathscr{C}\mathscr{A}^{k-3}\mathscr{B} & \mathscr{C}\mathscr{A}^{k-4}\mathscr{B} & \cdots & \mathscr{D} \end{pmatrix}.$$

Finally the covariance matrix for one column in $H_h^{\mathrm{s}}E_{\mathrm{f}}$ is needed.

$$P_h \triangleq \mathrm{Cov}(H_h^{\mathrm{s}}e_{\mathrm{f}}) = H_h^{\mathrm{s}}(I_h \otimes R)(H_h^{\mathrm{s}})^{\mathrm{T}}.$$

The basic assumptions needed are listed below. They are very standard in system identification.

(A,C) is observable. (O)

$(A,[B \quad K])$ is controllable. (C)

The input u is quasi-*stationary*. (S)

The transfer function from e to y has all zeros strictly inside the unit circle. (Z)

The input u and noise e is jointly quasi-stationary and uncorrelated. (U)

Assumption (S) ensures that the limits for time averages involving u exists (Ljung, 1999, Definition 2.1). For these limits the notation $\bar{E}$ (2) will be used, it reduces to E in pure stationary stochastic cases and $\lim_{j\to\infty}(1/j)\sum_{k=1}^{j}$ in pure deterministic cases.

$$\bar{E}((\bullet)) \triangleq \lim_{j\to\infty} \frac{1}{j}\sum_{k=1}^{j} E((\bullet)_k). \tag{2}$$

Notice that uncorrelated in assumptions (U) involves a quasi-stationary signal and is then defined by (3) and holds for systems operating in open loop.

$$\bar{E}(u_{k+\tau}e_k^{\mathrm{T}}) = 0 \quad \forall\tau. \tag{3}$$

4. Estimating Γ_h, H_h^{d} and P_h

The overall estimation method can be outlined in three steps as follows: First, use the signal and parameter matrices to establish a linear regression model. Second, estimate a sufficient number of parameter matrices. The choice in this paper is Γ_h, H_h^{d} and P_h. Third, based on these matrices extract the basic parameters in the model (Definition 1). The third step is postponed until Section 6.

The first matrix equation (4) is derived directly from the model in Definition 1.

$$Y_{\mathrm{f}} = \Gamma_h X_{\mathrm{f}} + H_h^{\mathrm{d}}U_{\mathrm{f}} + H_h^{\mathrm{s}}E_{\mathrm{f}}. \tag{4}$$

Unfortunately Γ_h cannot be estimated from this model because X_{f} is not measurable. Therefore X_{f} is related to measurable signals, i.e. input and output as follows:

$$\begin{aligned} x_k &= Ax_{k-1} + Bu_{k-1} + Ke_{k-1} \Rightarrow \\ y_{k-1} &= Cx_{k-1} + Du_{k-1} + e_{k-1}, \\ x_k &= (A-KC)x_{k-1} + Ky_{k-1} + (B-KD)u_{k-1} \\ &= (A-KC)^2x_{k-2} + (A-KC)(Ky_{k-2} \\ &\quad + (B-KD)u_{k-2}) + Ky_{k-1} + (B-KD)u_{k-1} \\ &\;\;\vdots \\ &= (A-KC)^i x_{k-i} + (A-KC)^{i-1}(Ky_{k-i} \\ &\quad + (B-KD)u_{k-i}) + \cdots + (A-KC)(Ky_{k-2} \\ &\quad + (B-KD)u_{k-2}) + Ky_{k-1} + (B-KD)u_{k-1}. \end{aligned}$$

If the terms with y are taken first followed by terms with u and the equation is collected for $k = i \ldots i+j-1$ the second matrix equation below is obtained.

$$X_{\mathrm{f}} = L_y Y_{\mathrm{p}} + L_u U_{\mathrm{p}} + L_x X_{\mathrm{p}}, \tag{5}$$

$$L_y \triangleq \mathscr{C}_i(A-KC,K), \qquad L_u \triangleq \mathscr{C}_i(A-KC,B-KD),$$
$$L_x \triangleq (A-KC)^i. \tag{6}$$

Inserting (5) in (4) gives (7) which can be written in a more regression type of way (8) by introducing Definition 4.

$$Y_{\mathrm{f}} = \Gamma_h(L_y Y_{\mathrm{p}} + L_u U_{\mathrm{p}} + L_x X_{\mathrm{p}}) + H_h^{\mathrm{d}}U_{\mathrm{f}} + H_h^{\mathrm{s}}E_{\mathrm{f}}. \tag{7}$$

Definition 4 (*Regression parameters and regressors*).

$$\Theta_{\mathrm{p}} \triangleq \Gamma_h[L_y \quad L_u], \qquad \Theta_{\mathrm{f}} \triangleq H_h^{\mathrm{d}}, \qquad \Theta \triangleq [\Theta_{\mathrm{p}} \quad \Theta_{\mathrm{f}}],$$
$$W_{\mathrm{p}} \triangleq \begin{bmatrix} Y_{\mathrm{p}} \\ U_{\mathrm{p}} \end{bmatrix}, \qquad Z \triangleq \begin{bmatrix} W_{\mathrm{p}} \\ U_{\mathrm{f}} \end{bmatrix}.$$

Note that (8) is a LS regression model in the sense that the residuals, columns in $H_h^{\mathrm{s}}E_{\mathrm{f}}$, is uncorrelated with the regressors, columns in Z and X_{p}, due to assumption (U) and e_k being white noise.

$$\begin{aligned} Y_{\mathrm{f}} &= [\Gamma_h L_y \quad \Gamma_h L_u \; H_h^{\mathrm{d}}] \begin{bmatrix} Y_{\mathrm{p}} \\ U_{\mathrm{p}} \\ U_{\mathrm{f}} \end{bmatrix} + \Gamma_h L_x X_{\mathrm{p}} + H_h^{\mathrm{s}}E_{\mathrm{f}} \\ &= [\Theta_{\mathrm{p}} \quad \Theta_{\mathrm{f}}] \begin{bmatrix} W_{\mathrm{p}} \\ U_{\mathrm{f}} \end{bmatrix} + \Gamma_h L_x X_{\mathrm{p}} + H_h^{\mathrm{s}}E_{\mathrm{f}} \\ &= \Theta Z + \Gamma_h L_x X_{\mathrm{p}} + H_h^{\mathrm{s}}E_{\mathrm{f}}. \end{aligned} \tag{8}$$

If the term with X_p in (8) is overlooked the regression model (9) appears and the LS estimate (10) is obtained if Z has full row rank.

$$Y_f = \Theta Z + V, \tag{9}$$

$$\hat{\Theta} = Y_f Z^T (ZZ^T)^{-1}. \tag{10}$$

Section 5 shows that this can give a useful estimate of $\hat{\Theta}$. Assume therefore for a while that Θ is known and that the residual $v = H_h^s e_f$. Under this assumption H_h^d is directly given by Θ_f (11) and the covariance P_h for $H_h^s e_f$ is estimated by (12).

$$H_h^d = \Theta_f, \tag{11}$$

$$\hat{P}_h = \frac{1}{j} V V^T. \tag{12}$$

Finally Γ_h is found by the SVD step below which is similar to the use of SVD in other subspace approaches.

Theorem 1 (Γ_h from Θ). *Under assumption* (O), (C), $h \geq n$, $i \geq n$, $W_1 \in \mathbb{R}^{hl \times hl}$ *and nonsingular and* $W_2 \in \mathbb{R}^{i(l+m) \times r}$, $r \geq i(l+m)$ *and full rank,* Γ_h *is given by* (13) *where* T *is a similarity transformation and* U_1 *is calculated by the singular value decomposition* (14).

$$\Gamma_h = W_1^{-1} U_1 T, \quad |T| \neq 0, \ T \in \mathbb{R}^{n \times n}, \tag{13}$$

$$W_1 \Theta_p W_2 = USV^T = [U_1 \ U_2] \begin{bmatrix} S_1 & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{0}} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \quad S_1 \in \mathbb{R}^{n \times n}. \tag{14}$$

Remark 1.1. Notice that Γ_h is not unique but dependent on the users choice of T. If the state space basis x_I corresponds to using $T = I$ then the state space basis $x = T^{-1} x_I$ corresponds to the choice T. The choice of T will affect the estimates of parameters A, B, C, K and Γ_h discussed in the rest of the paper in a well-known way (Zhou, Doyle & Glover, 1996).

Remark 1.2. From practical experience the impact of T seems to be small, $T = I$, i.e. omitting T is therefore the simple choice used in this paper.

Remark 1.3. Practical experience also shows that $W_2 = (W_p W_p^T)^{1/2}$ gives less uncertainty than $W_2 = I$ so the choice here is the former and $W_1 = I$ as W_1 seems to have only little impact.

Proof. Γ_h has full column rank n under (O) if $h \geq n$. In Appendix A it is proven that $[L_y \ L_u]$ has full row rank n under (C) if $i \geq n$, then due to $\Theta_p = \Gamma_h [L_y \ L_u]$ (Definition 4) Γ_h and Θ_p have the same image (left space). Consequently $\text{im}(\Gamma_h) = \text{im}(\Theta_p) = \text{im}(W_1^{-1} U_1)$ under the assumptions on W_1 and W_2 which completes the proof. □

Based on measurement the system matrices are calculated as follows.

Definition 5 (Calculating $\hat{\Gamma}_h$, $\hat{H}_h^d$ and $\hat{P}_h$). Based on input/output data $\hat{\Gamma}_h$, $\hat{H}_h^d$ and $\hat{P}_h$ are obtained by substituting $\hat{\Theta}$ for Θ and $Y_f - \hat{\Theta} Z$ for V in (11)–(14).

5. Consistency analysis for $\hat{\Theta}$, $\hat{\Gamma}_h$, $(\hat{\Gamma}_h^{\perp})^T \hat{H}_h^d$, $\hat{H}_h^d$ and $\hat{P}_h$

Inserting (8) in (10) gives (15) which allows the main Theorems 2 and 3.

$$\begin{aligned} \hat{\Theta} &= Y_f Z^T (ZZ^T)^{-1} \\ &= (\Theta Z + \Gamma_h L_x X_p + H_h^s E_f) Z^T (ZZ^T)^{-1} \\ &= \Theta + \Gamma_h L_x X_p Z^T (ZZ^T)^{-1} + H_h^s E_f Z^T (ZZ^T)^{-1}. \end{aligned} \tag{15}$$

Theorem 2 (Limit for $\hat{\Theta}$). *Assuming* (S), *the input persistently exciting of order* $i + h$ *and* (U) *then*

$$\begin{aligned} &\hat{\Theta} \to \Theta + \Gamma_h L_x \Delta \quad \text{for } j \to \infty \quad \text{(wp1)} \\ &\Leftrightarrow [\hat{\Theta}_p \quad \hat{\Theta}_f] \to [\Theta_p \quad \Theta_f] + \Gamma_h L_x [\Delta_p \quad \Delta_f] \\ &\text{for } j \to \infty \quad \text{(wp1)}, \end{aligned} \tag{16}$$

where

$$\begin{aligned} &X_p Z^T (ZZ^T)^{-1} \to \Delta = [\Delta_p \ \Delta_f] \quad \text{for } j \to \infty \quad \text{(wp1)}, \\ &\Delta \in \mathbb{R}^{n \times i(l+m)+hm}, \quad \Delta_p \in \mathbb{R}^{n \times i(l+m)}, \quad \Delta_f \in \mathbb{R}^{n \times hm}. \end{aligned}$$

Proof. Due to (U) e_f and z are uncorrelated so $\bar{E}(e_f z^T) = \underline{\underline{0}}$ and $\bar{E}(zz^T)$ exist due to (S) and is nonsingular because of persistent excitation (Söderström & Stoica, 1989). The limit for the third term in (15) is then $\underline{\underline{0}}$ by (17) and the limit of the second term exists due to (S) which completes the proof.

$$\begin{aligned} E_f Z^T (ZZ^T)^{-1} &= \frac{1}{j} E_f Z^T \left(\frac{1}{j} ZZ^T \right)^{-1} \\ &= \frac{1}{j} \sum_{k=1}^{j} e_f(k) z(k)^T \left(\frac{1}{j} \sum_{k=1}^{j} z(k) z(k)^T \right)^{-1} \\ &\to \bar{E}(e_f z^T) \bar{E}(zz^T)^{-1} = \underline{\underline{0}} \\ &\quad \text{for } j \to \infty \quad \text{(wp1)}. \quad \square \end{aligned} \tag{17}$$

The limits below are necessary for the model parameter estimation step in Section 6. The orthogonal compliment $\hat{\Gamma}_h^{\perp}$ in the theorem is found as U_2 in the SVD of Theorem 1 because $W_1 = I$.

Theorem 3 (Consistency conditions for $\hat{\Gamma}_h$, $(\hat{\Gamma}_h^{\perp})^T \hat{H}_h^d$, $\hat{H}_h^d$ and $\hat{P}_h$). *Assuming* $h \geq n$, *the input persistently exciting of order* $i + h$ *and all basic assumptions* (O), (C), (S), (Z) *and* (U) *then* $\hat{\Gamma}_h$ *and* $(\hat{\Gamma}_h^{\perp})^T \hat{H}_h^d$ *are consistent for some limited* i_s *and* $i \geq i_s \geq n$ (18)–(19). $\hat{H}_h^d$ *and* $\hat{P}_h$ *are asymp-*

totically consistent (20)–(21).

$$\hat{\Gamma}_h \to \Gamma_h \quad \textit{for } j \to \infty,\ i \geq i_s \quad \text{(wp1)}, \tag{18}$$

$$(\hat{\Gamma}_h^{\perp})^{\mathrm{T}} \hat{H}_h^{\mathrm{d}} \to (\Gamma_h^{\perp})^{\mathrm{T}} H_h^{\mathrm{d}} \quad \textit{for } j \to \infty,\ i \geq i_s \quad \text{(wp1)}, \tag{19}$$

$$\hat{H}_h^{\mathrm{d}} \to H_h^{\mathrm{d}} \quad \textit{for } j \to \infty,\ i \to \infty \quad \text{(wp1)}, \tag{20}$$

$$\hat{P}_h \to P_h \quad \textit{for } j \to \infty,\ i \to \infty \quad \text{(wp1)}. \tag{21}$$

Remark 3.1. Notice that for (20) and (21) the assumption is that the input is persistently exciting of any order.

Remark 3.2. In (20) and (21) j must increase sufficiently much faster than i in order for (17) to hold, see e.g. Peternell et al. (1996).

Proof. In (16) the bias term (22) includes $L_x = (A - KC)^i$ which decreases with i because the eigenvalues of $(A - KC)$ equals the zeros in the transfer function from e to y which are strictly inside the unit circle due to (Z). This proves (20) and (21).

$$\Theta_{\mathrm{b}} \triangleq \Gamma_h L_x \Delta = \Gamma_h (A - KC)^i \Delta \to \underline{\underline{0}} \quad \text{for } i \to \infty. \tag{22}$$

From Theorem 2 and Definition 4 it follows that the limit for Θ_{p} is left multiplied by Γ_h(23). If the matrix $L = [[L_y\ L_u] + L_x \Delta_{\mathrm{p}}]$ has full rank then $\hat{\Gamma}_h$ is consistent according to Theorem 1. Theorem A.1 in Appendix A guarantees that $[L_y\ L_u]$ has full rank under (C) and $L_x = (A - KC)^i$ can be made arbitrarily small by increasing i. Because of continuity there exists an i where the matrix L does not lose rank and finally singular values and the space spanned by the left singular vectors are continuous (Bauer, 1998, p. 160; Harville, 1997, p. 564) which proves (18)

$$\hat{\Theta}_{\mathrm{p}} \to \Theta_{\mathrm{p}} + \Gamma_h L_x \Delta_{\mathrm{p}} = \Gamma_h [[L_y\ L_u] + L_x \Delta_{\mathrm{p}}] = \Gamma_h L \quad \text{for } j \to \infty \quad \text{(wp1)}. \tag{23}$$

From the above and Theorem 2 follows that $\hat{\Gamma}_h^{\perp} \to \Gamma_h^{\perp}$ and $\hat{H}_h^{\mathrm{d}} = \hat{\Theta}_{\mathrm{f}} \to H_h^{\mathrm{d}} + \Gamma_h L_x \Delta_{\mathrm{f}}$ when $j \to \infty$ which will cancel the bias part of $\hat{H}_h^{\mathrm{d}}$ as $(\Gamma_h^{\perp})^{\mathrm{T}} \Gamma_h = \underline{\underline{0}}$ which proves (19). □

Theorem 3 is the main result of this paper for the following reasons. The sufficient conditions for consistency of $\hat{\Gamma}_h$ given in Jansson and Wahlberg (1998) for colored persistent exciting input is that (A, B) is controllable which excludes the large and important class of BJ-type models. Theorem 3 only requires (C), i.e. $(A, [B\ K])$ to be controllable and consequently does not have this limitation. On the other hand consistency is only guaranteed for i not less than some limited but unspecified $i_s \geq n$ which is in accordance with the counterexample in Jansson and Wahlberg (1998) where $i = n$. Practise shows that i_s does not need to be large, often the minimal n is sufficient as in the BJ examples in Section 7. By including an extra assumption Chui (1997), Theorem 5.18 obtains for his approach a more strict sufficient condition corresponding to $i \geq 3n$, i.e. $i_s = 3n$.

As indicated in Theorem 3 there will be bias on $\hat{P}_h$ for limited i. This bias decreases fast with i as stated in the theorem below where the proof is omitted.

Theorem 4 (Limit for $\hat{P}_h$). *Under the assumptions of Theorem* 2 *the limit for* $\hat{P}_h$ *is*

$$\hat{P}_h \to P_h + \Gamma_h L_x P_{\tilde{x}_{\mathrm{p}}} L_x^{\mathrm{T}} \Gamma_h^{\mathrm{T}} \quad \textit{for } j \to \infty \quad \text{(wp1)}, \tag{24}$$

$$P_{\tilde{x}_{\mathrm{p}}} = \bar{E}(x_{\mathrm{p}} x_{\mathrm{p}}^{\mathrm{T}}) - \bar{E}(x_{\mathrm{p}} z^{\mathrm{T}}) \bar{E}(z z^{\mathrm{T}})^{-1} \bar{E}(z x_{\mathrm{p}}^{\mathrm{T}}). \tag{25}$$

Remark 4.1. The limit for $\hat{P}_h$ for limited i (24) includes $P_{\tilde{x}_{\mathrm{p}}}$ which is interpreted as the covariance for the estimation error $\tilde{x}_{\mathrm{p}} = x_{\mathrm{p}} - \hat{x}_{\mathrm{p}}|z$ which also decreases with i. Consequently the convergence for $\hat{P}_h$ with respect to i is fast due to the three factors $L_x P_{\tilde{x}_{\mathrm{p}}} L_x^{\mathrm{T}}$ all decreasing with i.

6. Estimation and consistency for model parameters

After having estimated the system matrices, e.g. Γ_h the model parameters must be estimated. For this, the literature holds a variety of methods. However, no method seems to be superior. The principle for methods, which avoids the use of an estimated state sequence, is to solve the following, or similar, equations for the model parameters. The right-hand sides are simply the functional relationship given in Definition 3. Left multiplying with $(\hat{\Gamma}_h^{\perp})^{\mathrm{T}}$ in the middle equation is done to avoid bias as explained in Theorem 3.

$$\hat{\Gamma}_h = \Gamma_h(A, C), \qquad (\hat{\Gamma}_h^{\perp})^{\mathrm{T}} \hat{H}_h^{\mathrm{d}} = (\hat{\Gamma}_h^{\perp})^{\mathrm{T}} H_h^{\mathrm{d}}(A, B, C, D),$$

$$\hat{P}_h = P_h(A, C, K, R). \tag{26}$$

As these equations are over-determined there are many solutions. A simple method is shown below.

Theorem 5 (Estimated model parameters from estimated system matrices). *Let model parameters be estimated by* (29)–(33) *where a* MATLAB *like notation is used and* † *denotes the More–Penrose pseudoinverse. Assume* $h \geq n + 1$, *the input persistently exciting of order* $i + h$ *and all basic assumptions* (O), (C), (S), (Z) *and* (U) *then*

$$\hat{A}, \hat{B}, \hat{D}, \textit{ and } \hat{C} \textit{ are consistent for some } i_s \textit{ and } i \geq i_s \tag{27}$$

$$\hat{K} \textit{ and } \hat{R} \textit{ are assymtotically consistent, i.e. for } i \to \infty \tag{28}$$

$$\hat{C} = \hat{\Gamma}_h(1\!:\!l, 1\!:\!n), \tag{29}$$

$$\hat{A} = (\hat{\Gamma}_h^u)^{\dagger} \hat{\Gamma}_h^l, \qquad \hat{\Gamma}_h^u \triangleq \hat{\Gamma}_h(1\!:\!(h-1)l, :),$$

$$\hat{\Gamma}_h^l \triangleq \hat{\Gamma}_h(l+1\!:\!hl, :), \tag{30}$$

$$\begin{bmatrix}\hat{B}\\ \hat{D}\end{bmatrix} = \arg\min_{B,D} |(\hat{\Gamma}_h^{\perp})^{\mathrm{T}}\hat{H}_h^{\mathrm{d}} - (\hat{\Gamma}_h^{\perp})^{\mathrm{T}}H_h^{\mathrm{d}}(\hat{A},B,\hat{C},D)|_2^2, \tag{31}$$

$$\hat{R} = \hat{P}_h(1:l,1:l), \tag{32}$$

$$\hat{K} = (\hat{\Gamma}_h^{u})^{\dagger}\hat{P}_h(l+1:hl,1:l)\hat{R}^{-1}. \tag{33}$$

Remark 5.1. For a limited number of samples there is a risk that the estimated model parameters result in an unstable predictor. In this case a stable $\hat{K}$ and $\hat{R}$ are obtained from a stationary Kalman filter based on the first estimates. This additional step is similar to mirroring unstable zeros in a noise transfer function and will leave estimated models with a stable predictor unchanged.

Remark 5.2. A much simpler but biased alternative to (31) is

$$\hat{D} = \hat{H}_h(1:l,\ 1:m), \qquad \hat{B} = (\hat{\Gamma}_h^{u})^{\dagger}\hat{H}_h(l+1:hl,\ 1:m).$$

Remark 5.3. The bias on the stochastic part for limited i seems to be a unresolved problem.

Proof. According to Theorem 3 and the continuity properties of More–Penrose pseudoinverse all the matrices $\hat{\Gamma}_h, (\hat{\Gamma}_h^{u})^{\dagger}, \hat{\Gamma}_h^{l}$ and $(\hat{\Gamma}_h^{\perp})^{\mathrm{T}}\hat{H}_h^{\mathrm{d}}$ tends to there true values as $j\to\infty$ for fixed $i \geq i_s$ whereas the matrix $\hat{P}_h$ only converges to the true value if also $i\to\infty$. Then Eqs. (29)–(33) only must be verified for true values to prove (27)–(28).

Eq. (29) follows directly from Definition 3. If $h \geq n+1$ then $\Gamma_h^{u} = \Gamma_{h-1}$ has full column rank n under (O) and then (30) is proved below by using the shift property of Γ_h.

$$\Gamma_h^{u}A = \Gamma_h^{l} \ \Rightarrow\ A = \Gamma_h^{u\dagger}\Gamma_h^{l}.$$

The minimization in (31) is a LS problem because the squared term is linear in B, D, see Verhaegen (1994), Westwick and Verhaegen (1996) and Bauer (1998, p. 146) for details on a similar method based on a alternative estimate of H_h^{d}. Clearly B, D is a solution to (31) in the limit. Now it only remains to show that (31) has a unique solution (in the limit) which can be proved Bauer (1998, p. 147). This completes the proof of (27).

Eq. (32) follows directly from Definition 3 and (33) is proven in a similar way as (30) by observing that

$$P_h(l+1;\ hl,\ 1:l) = \begin{bmatrix} CKR \\ CAKR \\ \vdots \\ CA^{h-2}KR \end{bmatrix} = \Gamma_{h-1}KR.$$

which ends the proof of (28). □

7. Numerical examples

Below a few examples illustrate the theorems above. First the possible but unlikely lag of consistency for small i is demonstrated by means of the counterexample constructed and described in Jansson and Wahlberg (1998). Then it is shown that raising i gives consistency for the deterministic part also for the counterexample. An example from the widely used class of BJ systems is also included to show consistency and bias for the deterministic and stochastic parts, respectively.

The counterexample (34) is SISO with a low-pass deterministic part with a double pole at 0.9184. It is controllable from the input only. The stochastic part has the same poles as the deterministic and zeros at $0.9791 \pm 0.1045i$ with an absolute value of 0.9847. These zeros are very close to the unit circle which is important because these are the poles in the observer, i.e. eigenvalues in $A-KC$. The input is a fourth-order moving average process of high-pass type which in practice would never be chosen for a low-pass system. The S/N is approximately 0.5 which is very low. Thus, both system and experimental conditions are really worst case.

$$y(k) = \frac{B(q)}{A(q)}u(k) + \frac{C(q)}{A(q)}e(k), \qquad \mathrm{E}(e(k)) = 0,$$

$$\mathrm{E}(e(k)e(l)) = \delta_{kl}\sigma_e^2. \tag{34}$$

In the counterexample $h=3$ and $i=2$ which are the smallest possible. For this case and for $h=i=3$ the parameters are estimated with an increasing number of samples. The method used is given by Definition 5, Remarks 1.2–1.3 and Theorem 5. The estimated state space parameters are converted to parameters in the transfer function model (34). The resulting parameter estimation errors are shown in Figs. 1 and 2 and Table 1. As shown in Jansson and Wahlberg (1998), the estimates do not converge to the correct parameters for $i=2$. The result of raising i to 3 is seemingly consistent for all dynamic parameters as illustrated in Fig. 2. For the deterministic part this is explained by Theorem 5. The figures in Table 1 for $i=3$ shows a relative error for the determinis-

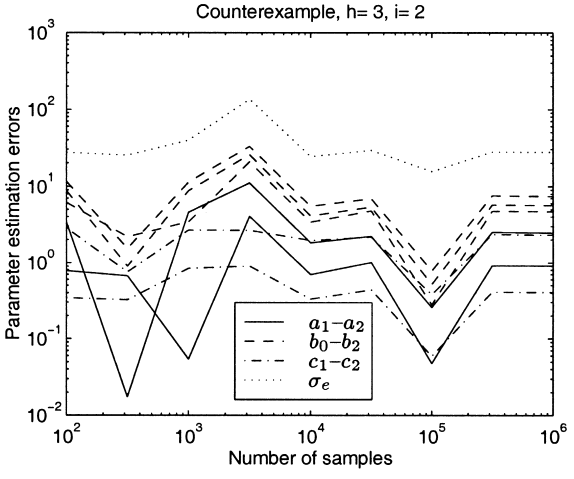

Fig. 1. Convergence for counter example with $h=3$ and $i=2$.

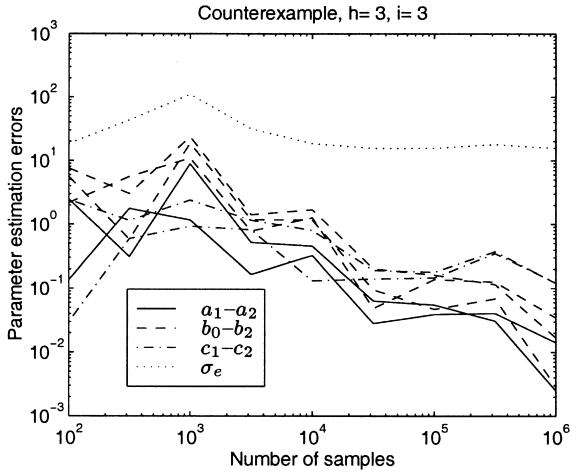

Fig. 2. Convergence for the modified counter example with $h = 3$ and $i = 3$.

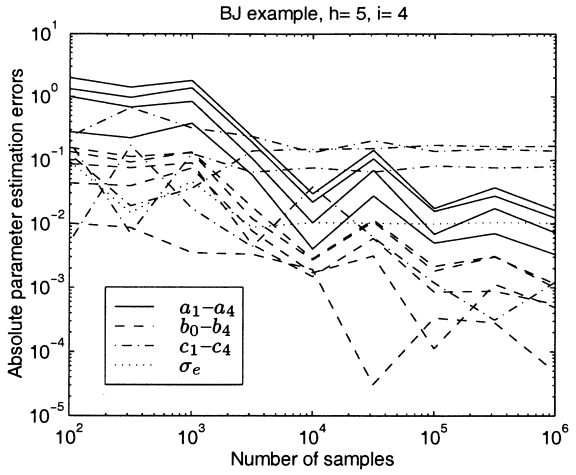

Fig. 3. Convergence for fourth-order BJ example with $h = 5$ and $i = 4$.

Table 1
Maximum relative error in transfer function coefficient estimates, if the correct parameter is zero then the absolute error is used. The number of samples are 10^6 in all cases. The deterministic part includes nominator and the common denominator. The stochastic part includes only the nominator

Example	h	i	Det.	Stoch.	σ_e
Counter	3	2	3.622	1.180	276.089
Counter	3	3	0.027	0.127	155.165
BJ	5	4	0.011	0.362	0.103
BJ	5	12	0.005	0.026	0.005

tic part at less than 3% whereas the figure is 13% for the dynamical stochastic part and huge for the white noise variance. Consequently, what seems to be consistency for c parameters in Fig. 2 is really a small bias as explained by Remark 4.1.

The BJ example (35) is a fourth-order SISO system with a low-pass deterministic and stochastic part. They are both discretizations of second-order continuous systems with natural frequency and damping 0.1 Hz and 0.5 for the deterministic part $B(q)/F(q)$ and 0.05 Hz and 0.7 for the stochastic part $C(q)/D(q)$. Furthermore there is a direct feed-through with gain 0.1. This system is of course not controllable from the input alone. The input is band-limited white noise with a cut off frequency of 0.125 Hz (Söderström & Stoica, 1987), Example 5.11). The S/N is approximately 1 which does not make the estimation too easy.

$$y(k) = \frac{B(q)}{F(q)}u(k) + \frac{C(q)}{D(q)}e(k), \qquad \mathrm{E}(e(k)) = 0,$$

$$\mathrm{E}(e(k)e(l)) = \delta_{kl}\sigma_e^2. \tag{35}$$

In Fig. 3 and Table 1 the results of using $h = 5$ and $i = 4$ are shown. Notice that this h and i are the smallest possible. Clearly, the deterministic part is consistent, the maximal relative error after 10^6 samples is 1% whereas it is 36% for the stochastic part which is biased.

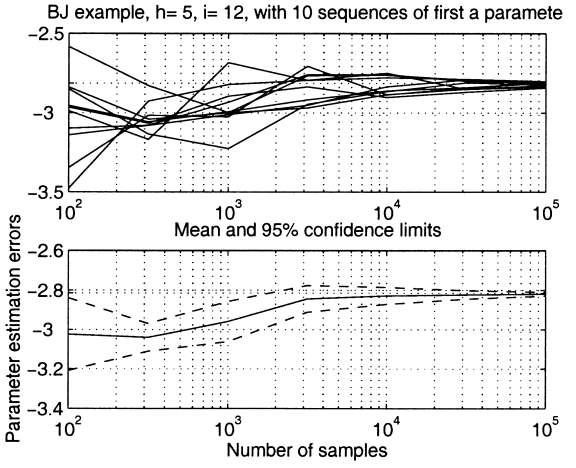

Fig. 4. Convergence for a_1 in the fourth-order BJ example with $h = 5$ and $i = 12$. The upper plot shows 10 sequences of estimates for a_1, the correct parameter $a_1 = -2.814$ is shown with an extra tick mark. The lower plot shows the corresponding averages with 95% confidence limits.

Another example using $h = 5$ and $i = 12$ are shown in Table 1 and Fig. 4. Here only the estimate of a_1 is shown but for 10 sequences of 10^6 samples. In general consistency does not imply unbiased estimators for a limited number of samples. This is exactly the case with this subspace method. The estimates are consistent but there is bias for at least some fixed number of samples. The picture for the other parameters is the same as for a_1.

8. Conclusion

The derivation of the subspace method is in this paper based on an LS regression model which is a simple alternative approach given new insight.

The particular method for estimating the stochastic parameters K and R is simple and new to the authors knowledge. Simulations not included here indicate that it is superior for BJ models.

However, the consistency analysis with focus on the extended observability matrix is the main contribution. Previously sufficient conditions have been given which apply to most practical identification problems. Unfortunately new published result shows that these conditions are not sufficient in all cases and then gives new stricter conditions which excludes many practical problems as, e.g. BJ type of models. In this paper the consistency conditions boil down to a sum of two matrices having full rank. It is shown in Theorem 3 that the first one has full rank under the general conditions whereas the second one can result in loss of rank. This supports the new result. The encouraging result in the paper then is that the second matrix can be made arbitrarily small by including a sufficient number of past data in the regression. Consequently we are back to the previous general conditions if the number of past data is chosen correctly. In this case the consistency is shown to extend to the complete deterministic part of the model.

Appendix A. Controllability conditions

Theorem A.1. $[L_y \ L_u]$ *defined in* (6) *has full row rank n if and only if* $(A, [B \ K])$ *is controllable* (*assumption* (C)).

Proof.

$$[L_y \quad L_u] = [\mathscr{C}_i(A - KC, K) \quad \mathscr{C}_i(A - KC, B - KD)]$$
$$= [(A - KC)^i K \quad \cdots$$
$$K(A - KC)^i(B - KD) \quad \cdots \quad (B - KD)].$$

Sorting the blocks by i retains the rank and gives the observability matrix

$$\mathscr{C}_i(A - KC, [B - KD \ K])$$
$$= [(A - KC)^i[B - KD \ K] \quad \cdots \quad [B - KD \ K]].$$

This matrix has full rank n if the system $(A - KC, [B - KD \ K])$ is controllable which is equivalent to the closed-loop system matrix (A.2) which can be assigned any eigenvalues by choosing the feedback F (Zhou et al., 1996, Section 3.2). Introduce the bijective mapping (A.1). Then the two systems $(A - KC, [B - KD \ K], F)$ and $(A, [B \ K], F')$ have the same closed-loop system matrix (A.2) and (A.3), here the notation (system matrix, input matrix, feedback matrix) is used. Consequently if and only if a set of eigenvalues can be assigned to the first system by F it can also be assigned to the second by F' which completes the proof.

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \qquad F' = \begin{bmatrix} F'_1 \\ F'_2 \end{bmatrix},$$

$$F_2 = F'_2 + C + DF_1 \Leftrightarrow F'_2 = F_2 - C - DF_1,$$

$$F_1 = F'_1, \tag{A.1}$$

$$A - KC + [B - KD \quad K]\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{A.2}$$
$$= A - KC + (B - KD)F_1 + KF_2$$
$$= A + BF_1 + KF'_2$$
$$= A + [B \quad K]\begin{bmatrix} F'_1 \\ F'_2 \end{bmatrix}. \qquad \square \tag{A.3}$$

References

Bauer, D., Deistler, M., & Scherrer, W. (1997). The analysis of the asymptotic variance of subspace algorithms. In Y. Sawaragi, & S. Sagara (Eds.), *SYSID '97*, vol. 3, IFAC the International Federation of Automatic Control (pp. 1087–1091).

Bauer, D. (1998). *Some asymptotic theory for the estimation of linear systems using maximum likelihood methods or subspace algorithms.* Ph.D. thesis, Institute of Econometrics, Operations Research and System Theory, Vienna University of Technology.

Chui, N. L. C. (1997). *Subspace methods and informative experiments for system identification.* Ph.D. thesis, Pembroke College Cambridge. Obtained at http://www-control.eng.com.ac.uk/nlcc/report/thesis.ps.

Deistler, M., Peternell, K., & Scherrer, W. (1995). Consistency and relative efficiency of subspace methods. *Automatica*, *31*(12), 1865–1875.

Gustafsson, T. (1997). System identification using subspace-based instrumental variable methods. In Y. Sawaragi, & S. Sagara (Eds.), *SYSID '97*, vol. 3, IFAC, IFAC the International Federation of Automatic Control (pp. 1119–1124).

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective.* Berlin: Springer.

Jansson, M., & Wahlberg, Bo. (1996). A linear regression approach to state-space subspace system identification. *Signal Processing*, *52*(2), 103–129.

Jansson, M., & Wahlberg, Bo. (1998). On consistency of subspace methods for system identification. *Automatica*, *34*(12), 1507–1519.

Ljung, L. (1999). *System identification, theory for the user* (2nd ed.). Information and system sciences series. Englewood Cliffs, NJ: Prentice-Hall.

Ljung, L., & McKelvey, T. (1997). Interpretation of subspace methods: Consistency analysis. In Y. Sawaragi, & S. Sagara (Eds.), *SYSID '97*, vol. 3, IFAC, IFAC the International Federation of Automatic Control (pp. 1125–1129).

Ohsumi, A., Takashima, S., & Kameyama, K. (1997). Subspace-based identification of stochastic systems using innovation model. In Y. Sawaragi, & S. Sagara (Eds.), *SYSID '97*, vol. 3, IFAC the International Federation of Automatic Control (pp. 1113–1118).

Ottersten, B., & Viberg, M. (1994). A subspace based instrumental variable method for state-space system identification. In M. Blanke, & T. Söderström (Eds.), *SYSID '97*, vol. 2, IFAC, Danish Automation Society (pp. 139–144).

Peternell, K., Scherrer, W., & Deistler, M. (1996). Statistical analysis of novel subspace identification methods. *Signal Processing*, *52*, 161–177.

Peternell, K. (1995). *Identification of linear dynamic systems by subspace and realization-based algorithms*. Ph.D. thesis, Institute of Econometrics, Operations Research and System Theory, Vienna University of Technology.

Picci, G. (1997). Statistical properties of certain subspace identification methods. In Y. Sawaragi, & S. Sagara (Eds.), *SYSID '97*, vol. 3, IFAC, IFAC the International Federation of Automatic Control (pp. 1093–1099).

Söderström, T., & Stoica, P. (1989). *System identification*. International series in system and control engineering. New York: Prentice-Hall.

Sorelius, J., Söderström, T., Stoica, P., & Cedervall, M. (1997). Order estimation method for subspace-based system identification. In Y. Sawaragi, & S. Sagara (Eds.), *SYSID '97*, vol. 3, IFAC, IFAC the International Federation of Automatic Control (pp. 1131–1136).

van Overschee, P., & De Moor, B. (1996). *Subspace identification for linear systems*. Dordrecht: Kluwer Academic Publishers.

Verhaegen, M. (1993). Identification of the deterministic and stochastic part of mimo state space models under the presence of process and measurement noise. *European control conference* (pp. 1313–1318).

Verhaegen, M. (1994). Identification of the deterministic part of mimo state space models given in innovations form from input–output data. *Automatica*, *30*(1), 61–74.

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

Westwick, D., & Verhaegen, M. (1996). Identifying mimo wiener systems using subspace model identification methods. *Signal Processing*, *52*, 235–258.

Zhou, K., Doyle, J. C., & Glover, K. (1996). *Robust and optimal control*. Englewood Cliffs, NJ: Prentice-Hall.

Torben Knudsen was born in 1958 in Aarhus, Denmark. He received his MS from the Technical University of Denmark in 1983. The Ph.D. degree was obtained in 1990 at the Institute for Mathematical Modeling at the same university. In 1986 he was at Wind Energy and Atmospheric Physics Department at Risø National Laboratory working as a research assistant. Since 1987 he has been at Department of Control Engineering, Aalborg University, Denmark where he is now an associate professor. His main research areas are stochastic systems, state estimation, system identification and control. Concerning applications he has been working mostly with wind turbines and recently with physiological systems.