



## Subspace Identification

*Reducing Uncertainty on the Stochastic Part*

Knudsen, Torben

*Publication date:*  
2001

*Document Version*  
Også kaldet Forlagets PDF

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Knudsen, T. (2001). *Subspace Identification: Reducing Uncertainty on the Stochastic Part*.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# SUBSPACE IDENTIFICATION - REDUCING UNCERTAINTY ON THE STOCHASTIC PART

**Torben Knudsen**

*Aalborg University, Institute of Electronic Systems, Department of  
Control Engineering, Fredrik Bajers Vej 7C, DK-9220 Aalborg Ø,  
Denmark, e-mail, tk@control.auc.dk*

**Abstract:** Subspace identification algorithms are user friendly, numerical fast and stable and they provide a good consistent estimate of the deterministic part of a system. The weak point is the stochastic part. The uncertainty on this part is discussed below and methods to reduce it is derived.

**Keywords:** Parameter estimation; linear multivariable systems; subspace methods; state space models; bias; stochastic part; least-squares methods

## 1. INTRODUCTION

Compared to prediction error and maximum likelihood methods the merits of subspace methods are robust numerical algorithms which does not require iterative minimization. Subspace methods are also easy to use as the only model structure information needed is the system order. The drawbacks are lower performance in terms of efficiency and also consistency for the stochastic part.

Recently, consistency for the whole deterministic part has been established under weak conditions (Bauer and Jansson, 2000; Knudsen, 2001). Unfortunately, similar good methods for the stochastic part under these weak conditions are still missing.

The purpose of this paper is to discuss the above problems and suggest methods to reduce it whit out introducing iterative minimization or otherwise spoiling the merits of subspace methods. Below, the overall problem is first stated then the notation and some necessary basic assumptions are given. Estimation of the deterministic part are reviewed as this gives the basis for the following discussion of estimates for the stochastic part and the uncertainty reducing methods suggested. These new methods is compared to existing ones by simulation experiments. Finally a conclusion is 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 which there is methods for (Sorelius *et al.*, 1997).

*Definition 1.* (Innovation model).

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k + Ke_k \\y_k &= Cx_k + Du_k + e_k \\E(e_p e_q^T) &\triangleq R\delta_{pq}\end{aligned}$$

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 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 is  $N = i + h + j - 1$ .

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

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  $k = i$  respectively.

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

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

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.* (Matrices related to parameter). The extended observability matrix  $\Gamma_k$  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 matrix  $C_i$  is defined below where  $A$  and  $B$  represent system and input matrices respectively.

$$C_i(A, B) \triangleq (A^{i-1}B \ A^{i-2}B \ \cdots \ AB \ B)$$

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  $\mathcal{H}_k$ .

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

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

$$\mathcal{H}_k(A, B, C, D) \triangleq \begin{pmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{k-2}B & CA^{k-3}B & \cdots & D \end{pmatrix}$$

Finally the covariance matrix for one column in  $H_h^s E_f$  is needed.

$$P_h \triangleq \text{Cov}(H_h^s e_f) = H_h^s (I_h \otimes R) (H_h^s)^T$$

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

- (O)  $(A, C)$  is observable
- (C)  $(A, [B \ K])$  is controllable
- (S) The input  $u$  is quasi-stationary
- (Z) The transfer function from  $e$  to  $y$  has all zeros strictly inside the unit circle
- (U) The input  $u$  and noise  $e$  is jointly quasi-stationary and uncorrelated

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

$$\bar{E}((\bullet)) \triangleq \lim_{j \rightarrow \infty} \frac{1}{j} \sum_{k=1}^j E((\bullet)_k) \quad (1)$$

Notice that uncorrelated in assumptions (U) involves a quasi-stationary signal and is then defined by (2) and holds for systems operating in open loop.

$$\bar{E}(u_{k+\tau} e_k^T) = 0 \quad \forall \tau \quad (2)$$

### 4. CONSISTENT ESTIMATE OF THE DETERMINISTIC PART

The necessary basis for analyzing the estimate of the stochastic part follows below. It is rather brief, for details and proofs please refer to (Knudsen, 2001). There are different theoretical frameworks for subspace identification. However, all frameworks have the focus on the deterministic part in common.

The overall estimation method chosen here 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  $\Gamma_h$ ,  $H_h^d$  and  $P_h$ . Third, based on these matrices extract the basic parameters in the model (definition 1).

#### 4.1 Linear regression model

The first matrix equation (3) is derived directly from the model in definition 1.

$$Y_f = \Gamma_h X_f + H_h^d U_f + H_h^s E_f \quad (3)$$

Unfortunately  $\Gamma_h$  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 (Knudsen, 2001).

$$X_f = L_y Y_p + L_u U_p + L_x X_p, \quad (4)$$

$$L_y \triangleq C_i(A - KC, K),$$

$$L_u \triangleq C_i(A - KC, B - KD), \quad (5)$$

$$L_x \triangleq (A - KC)^i$$

Inserting (4) in (3) gives (6) which can be written in a more regression type of way (7) by introducing definition 4.

$$Y_f = \Gamma_h(L_y Y_p + L_u U_p + L_x X_p) + H_h^d U_f + H_h^s E_f \quad (6)$$

*Definition 4.* (LS parameters and regressors).

$$\Theta_p \triangleq \Gamma_h [L_y \ L_u], \quad \Theta_f \triangleq H_h^d, \quad \Theta \triangleq [\Theta_p \ \Theta_f]$$

$$W_p \triangleq \begin{bmatrix} Y_p \\ U_p \end{bmatrix}, \quad Z \triangleq \begin{bmatrix} W_p \\ U_f \end{bmatrix}$$

Notice that (7) is a LS regression model in the sense that the residuals, columns in  $H_h^s E_f$ , is uncorrelated with the regressors, columns in  $Z$  and  $X_p$ , due to assumption (U) and  $e_k$  being white noise.

$$\begin{aligned} Y_f &= [\Gamma_h L_y \ \Gamma_h L_u \ H_h^d] \begin{bmatrix} Y_p \\ U_p \\ U_f \end{bmatrix} + \Gamma_h L_x X_p + H_h^s E_f \\ &= [\Theta_p \ \Theta_f] \begin{bmatrix} W_p \\ U_f \end{bmatrix} + \Gamma_h L_x X_p + H_h^s E_f \\ &= \Theta Z + \Gamma_h L_x X_p + H_h^s E_f \end{aligned} \quad (7)$$

Introduce the LS estimate (8) where  $Z$  must have full row rank.

$$\hat{\Theta} = Y_f Z^T (Z Z^T)^{-1} \quad (8)$$

Inserting (7) in (8) gives (9) from which the limit in theorem 1 can be derived.

$$\begin{aligned} \hat{\Theta} &= Y_f Z^T (Z Z^T)^{-1} \\ &= (\Theta Z + \Gamma_h L_x X_p + H_h^s E_f) Z^T (Z Z^T)^{-1} \\ &= \Theta + \Gamma_h L_x X_p Z^T (Z Z^T)^{-1} \\ &\quad + H_h^s E_f Z^T (Z Z^T)^{-1} \end{aligned} \quad (9)$$

*Theorem 1.* (Limit for  $\hat{\Theta}$ ). Assuming (S), the input persistently exciting of order  $i + h$  and (U) then

$$\lim_{j \rightarrow \infty} \text{wp1} \hat{\Theta} = \Theta + \Gamma_h L_x \Delta \Leftrightarrow \quad (10)$$

$$\lim_{j \rightarrow \infty} \text{wp1} \begin{bmatrix} \hat{\Theta}_p & \hat{\Theta}_f \end{bmatrix} = [\Theta_p \ \Theta_f] + \Gamma_h L_x [\Delta_p \ \Delta_f] \quad (11)$$

where

$$\lim_{j \rightarrow \infty} \text{wp1} X_p Z^T (Z Z^T)^{-1} = \Delta = [\Delta_p \ \Delta_f] \quad (12)$$

$$\begin{aligned} \Delta &\in \mathbb{R}^{n \times i(l+m) + hm}, \\ \Delta_p &\in \mathbb{R}^{n \times i(l+m)}, \quad \Delta_f \in \mathbb{R}^{n \times hm} \end{aligned}$$

## 4.2 Estimate parameter matrices

The key observations now are:  $\Theta_p$  is pre-multiplied by  $\Gamma_h$  and so is the bias in (10) i.e.  $\hat{\Theta}_p \rightarrow \Gamma_h L$ . Based on these observations it turns out that  $\Gamma_h$  can be consistently estimated under generic conditions using SVD (13)-(15). The bias on  $\hat{H}_h^d$  can be cancelled by projecting onto  $\hat{\Gamma}_h^\perp$  (18) the orthogonal complement to  $\hat{\Gamma}_h$  which is found as  $U_2$  (13).

*Theorem 2.* ( $\hat{\Gamma}_h$  from  $\hat{\Theta}_p$ ). 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  $i \geq i_s \geq n$  (16)–(18).

$$\begin{aligned} W_1 \hat{\Theta}_p W_2 &= U S V^T \\ &= [U_1 \ U_2] \begin{bmatrix} S_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \quad S_1 \in \mathbb{R}^{n \times n} \end{aligned} \quad (13)$$

$$T = I, \quad W_1 = I, \quad W_2 = (W_p W_p^T)^{\frac{1}{2}} \quad (14)$$

$$\hat{\Gamma}_h = W_1^{-1} U_1 T, \quad |T| \neq 0, \quad T \in \mathbb{R}^{n \times n} \quad (15)$$

$$\lim_{j \rightarrow \infty} \text{wp1} \hat{\Gamma}_h = \Gamma_h, \quad i \geq i_s \quad (16)$$

$$\hat{H}_h^d = \hat{\Theta}_f^d \quad (17)$$

$$\lim_{j \rightarrow \infty} \text{wp1} (\hat{\Gamma}_h^\perp)^T \hat{H}_h^d = (\Gamma_h^\perp)^T H_h^d, \quad i \geq i_s \quad (18)$$

*Remark 2.1.* Notice that  $\Gamma_h$  is not unique but dependent on the users choice where (14) works well.

## 4.3 Estimating model parameters

After having estimated the system matrices e.g.  $\Gamma_h$  the model parameters can be estimated by solving the following equations for the model parameters. The right hand sides are simply the functional relation giving in definition 3.

$$\hat{\Gamma}_h = \Gamma_h(A, C), \quad (19)$$

$$(\hat{\Gamma}_h^\perp)^T \hat{H}_h^d = (\hat{\Gamma}_h^\perp)^T H_h^d(A, B, C, D) \quad (20)$$

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

*Theorem 3.* (Estimating model parameters). Let model parameters be estimated by (22)–(24) where a MATLAB like notation is used and  $\dagger$  denotes the More-Penrose pseudo inverse. Assume  $h \geq n + 1$ , the input persistently exciting of order  $i + h$  and all basic assumptions (O), (C), (S), (Z) and (U) then

$$\begin{aligned} \hat{A}, \hat{B}, \hat{D}, \text{ and } \hat{C} &\text{ are consistent} \\ &\text{for some } i_s \text{ and } i \geq i_s \end{aligned} \quad (21)$$

$$\hat{C} = \hat{\Gamma}_h(1:l, 1:n) \quad (22)$$

$$\hat{A} = (\hat{\Gamma}_h^u)^\dagger \hat{\Gamma}_h^l, \quad (23)$$

$$\hat{\Gamma}_h^u \triangleq \hat{\Gamma}_h(1:(h-1)l, :),$$

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

$$\begin{aligned} \begin{bmatrix} \widehat{B} \\ \widehat{D} \end{bmatrix} &= \arg \min_{B,D} \\ &\left| (\widehat{\Gamma}_h^\perp)^T \widehat{H}_h^d - (\widehat{\Gamma}_h^\perp)^T H_h^d(\widehat{A}, B, \widehat{C}, D) \right|_2^2 \end{aligned} \quad (24)$$

*Remark 3.1.* The minimization in (24) is a LS problem because the squared term is linear in  $B, D$ , and (24) has a unique solution (in the limit) (Bauer, 1998, p. 147).

## 5. ESTIMATING THE STOCHASTIC PART

Following the estimation method in section 4 the residual covariance estimate  $\widehat{P}_h$  is calculated (25) and the parameters  $K, R$  is derived based on relation (26) and the limit for  $\widehat{P}_h$  derived below in theorem 4.

$$\begin{aligned} \widehat{P}_h &\triangleq \frac{1}{j - ((i+h)m + il)} \\ &\sum_{k=1}^j (y_f(k) - \widehat{\Theta}z(k))(y_f(k) - \widehat{\Theta}z(k))^T \\ &= \frac{1}{j - ((i+h)m + il)} (Y_f - \widehat{\Theta}Z)(Y_f - \widehat{\Theta}Z)^T \end{aligned} \quad (25)$$

$$\widehat{P}_h = P_h(\widehat{A}, \widehat{C}, K, R) \quad (26)$$

*Theorem 4.* (Limit for  $\widehat{P}_h$ ). Under the assumptions of theorem 1 the limit for  $\widehat{P}_h$  is

$$\lim_{j \rightarrow \infty} \text{wp1 } \widehat{P}_h = P_h + \Gamma_h L_x P_{\tilde{x}_p} L_x^T \Gamma_h^T, \quad (27)$$

$$P_{\tilde{x}_p} = \overline{\mathbb{E}}(x_p x_p^T) - \overline{\mathbb{E}}(x_p z^T) \overline{\mathbb{E}}(z z^T)^{-1} \overline{\mathbb{E}}(z x_p^T) \quad (28)$$

*Remark 4.1.* The limit for  $\widehat{P}_h$  (27) includes  $P_{\tilde{x}_p}$  which is interpreted as the covariance for the estimation error  $\tilde{x}_p = x_p - \hat{x}_p|z$  which also decreases with  $i$ . Consequently the convergence for  $\widehat{P}_h$  with respect to  $i$  is fast due to the three factors  $L_x P_{\tilde{x}_p} L_x^T$  all decreasing to  $\underline{0}$  with  $i$ .

*Proof.* According to (7) the residual is given by

$$\begin{aligned} v &= y_f - \hat{y}_f \\ &= \Theta z + \Gamma_h L_x x_p + H_h^s e_f - \widehat{\Theta} z \\ &= (\Theta - \widehat{\Theta})z + \Gamma_h L_x x_p + H_h^s e_f \end{aligned} \quad (29)$$

If the limiting residual is defined by (30) and the limit for  $\widehat{\Theta}$  (10) is inserted (31) is obtained.

$$\tilde{y}_f \triangleq \lim_{j \rightarrow \infty} \text{wp1 } v \quad (30)$$

$$\begin{aligned} \tilde{y}_f &= -\Gamma_h L_x \Delta z + \Gamma_h L_x x_p + H_h^s e_f \\ &= \Gamma_h L_x (x_p - \Delta z) + H_h^s e_f \end{aligned} \quad (31)$$

The definition (32) is introduced because if  $u$  is stochastic with mean zero then  $\Delta z$  is the optimal estimate of  $x_p$  given  $z$  which is denoted  $\hat{x}_p|z$ . The last equality below follows from (12) in theorem 1.

$$\tilde{x}_p \triangleq x_p - \Delta z = x_p - \overline{\mathbb{E}}(x_p z^T) \overline{\mathbb{E}}(z z^T)^{-1} z \quad (32)$$

Using (12) once again the following is obtained.

$$\begin{aligned} P_{\tilde{x}_p} &\triangleq \overline{\mathbb{E}}(\tilde{x}_p \tilde{x}_p^T) \\ &= \overline{\mathbb{E}}((x_p - \Delta z)(x_p - \Delta z)^T) \\ &= \overline{\mathbb{E}}(x_p x_p^T) - \overline{\mathbb{E}}(x_p z^T) \overline{\mathbb{E}}(z z^T)^{-1} \overline{\mathbb{E}}(z x_p^T) \end{aligned} \quad (33)$$

Now using that  $(x_p, z)$  and  $e_f$  are uncorrelated completes the proof as follows.

$$\begin{aligned} \overline{\mathbb{E}}(\tilde{y}_f \tilde{y}_f^T) &= \Gamma_h L_x \overline{\mathbb{E}}((x_p - \Delta z)(x_p - \Delta z)^T) \\ &\quad L_x^T \Gamma_h^T + H_h^s \mathbb{E}(e_f e_f^T) (H_h^s)^T \\ &= \Gamma_h L_x P_{\tilde{x}_p} L_x^T \Gamma_h^T + P_h \end{aligned} \quad (34)$$

Comparing the estimation problem for  $B, D$  and  $K, R$  reveals two important observations. The similarity is that the bias for  $\widehat{H}_h^d$  and  $\widehat{P}_h$  both lies in  $\text{im}(\Gamma_h)$ . However, the difference is that  $H_h^d(A, B, C, D)$  is linear in  $B, D$  while  $P_h(A, C, K, R)$  is not linear in  $K$ . This fact makes the estimation method for  $B, D$  (24) unusable for  $K, R$ .

A first estimation method for  $K, R$  which simply ignores the bias is given below. It is only based on the first block column in  $P_h$ .

$$\begin{aligned} P_h(1 : hl, 1 : l) &= \begin{bmatrix} R \\ CKR \\ CAKR \\ \vdots \\ CA^{h-2}KR \end{bmatrix} = \begin{bmatrix} R \\ \Gamma_h^u KR \end{bmatrix} \\ \widehat{R} &= \widehat{P}_h(1 : l, 1 : l) \end{aligned} \quad (35)$$

$$\widehat{K} = (\widehat{\Gamma}_h^u)^\dagger \widehat{P}_h(l+1 : hl, 1 : l) \widehat{R}^{-1} \quad (36)$$

The second method is based on Cholesky factoring of  $P_h$ . Notice that the Cholesky factor of a symmetric positive definite matrix is unique.

$$\begin{aligned} R &= GG^T, \\ P_h &= H_h^s (I_h \otimes R) (H_h^s)^T \\ &= H_h^s (I_h \otimes GG^T) (H_h^s)^T \\ &= H_h^s (I_h \otimes G) (H_h^s (I_h \otimes G))^T \\ &= \mathcal{H}_h(A, KG, C, G) \mathcal{H}_h(A, KG, C, G)^T, \\ \mathcal{H}_h(A, KG, C, G) &= \end{aligned}$$

$$\begin{pmatrix} G & 0 & \cdots & 0 \\ CKG & G & \cdots & 0 \\ CAKG & CKG & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CA^{h-2}KG & CA^{h-3}KG & \cdots & G \end{pmatrix}$$

With the Cholesky factorization below,  $K$  and  $R$  can be estimated by (37)–(40). The LS problem (38) is completely similar to (24) which is very convenient.

$$\widehat{P}_h = \widehat{\mathcal{H}}_h \widehat{\mathcal{H}}_h^T \quad (37)$$

$$\begin{aligned} \begin{bmatrix} \widehat{M} \\ \widehat{G} \end{bmatrix} &= \arg \min_{M,G} \\ &\left| (\widehat{\Gamma}_h^\perp)^T \widehat{\mathcal{H}}_h - (\widehat{\Gamma}_h^\perp)^T \mathcal{H}_h(\widehat{A}, M, \widehat{C}, G) \right|_2^2 \end{aligned} \quad (38)$$

$$\hat{R} = \hat{G}\hat{G}^T \quad (39)$$

$$\hat{K} = \hat{M}\hat{G}^{-1} \quad (40)$$

The bias in  $\hat{\mathcal{H}}_h$  does not lie exactly in  $\text{im}(\Gamma_h)$  but it probably has a large component there due to (27). Consequently the method does not cancel the bias term but it is lightly to reduce it.

A number of other approaches have been tried without success. Increasing  $i$  to decrease the bias term in (27) or using an estimate of  $L_y$  (5) to estimate  $K$  turned out to give poor performance in tests similar to those in section 6. The right hand side of (27) has also been rewritten into a linear function of  $R, KR, KKR^T$  and  $L_x P_{\bar{x}} L_x^T$ , unfortunately a unique solution to this equation could not be obtained for  $R$  and  $K$  in this way.

## 6. NUMERICAL EXAMPLES

The statistical performance of the methods is assessed by Monte Carlo simulation. Experimental conditions are 200 replications of 500 samples each. The S/N ratios is approximately 10 and a suitable excitation for the input  $u$  is used. To compare the results a state base independent representation is needed. All estimated state space models are therefore transformed into ARMAX transfer function representation (41).

$$\begin{aligned} A(q)y(t) &= B(q)u(t) + C(q)e(t), \\ \text{Var}(e) &\triangleq \sigma^2, \\ A(q) &\triangleq 1 + a_1q^{-1} + \dots + a_nq^{-n}, \\ B(q) &\triangleq b_0 + b_1q^{-1} + \dots + b_nq^{-n}, \\ C(q) &\triangleq 1 + c_1q^{-1} + \dots + c_nq^{-n}, \end{aligned} \quad (41)$$

A first, second and fourth order SISO system are tested. The first order system is used in (van Overschee and Moor, 1996, sect. 4.4.5) to illustrate these problems as it has a eigenvalue of  $A - KC$  at 0.9996 which is extremely close to the unit circle. Consequently,  $L_x = (A - KC)^i$  (5) decays slowly with  $i$  and so does the bias term for  $\hat{P}_h$  in (27). The two other, more well behaved systems are a second order ARMAX and a fourth order BJ system. The corresponding ARMAX representation are given in table 1.

All the system are tested with the methods in table 2. In this paper it is necessary to chose only one performance measure which is the rms for the parameter estimation errors (42).

$$\text{rms}_i = \frac{1}{M} \sum_{j=1}^M (\hat{\theta}_{ij} - \theta_{i0})^2 \quad (42)$$

For easy comparison the results show in figure 1–3 are normalized with the rms for the method tk-chol. For the first four subspace methods the choice of row numbers are  $h = i = 3, 4$  and  $6$  in the three test systems which are the smallest possible for the subid and n4sid methods. The fifth n4sid-auto method uses

automatic choice of  $h$  and  $i$ . Therefore it should have a potentially better performance. Consequently, only the first four subspace methods are directly comparable. The last two prediction error methods are included to give a lower limit for the obtainable rms.

Abb.	Method	Description
tk	subspace	(22)–(24), (35)–(36)
tk-chol	subspace	(22)–(24), (37)–(40)
subid	subspace	(van Overschee and Moor, 1996, p. 131)
n4sid	subspace	L. Ljung Ident Toolbox
n4sid-auto	subspace	L. Ljung Ident Toolbox
pem	PEM	L. Ljung Ident Toolbox
trftk	PEM	(Knudsen, 1996)

Table 2. Methods included in the test.

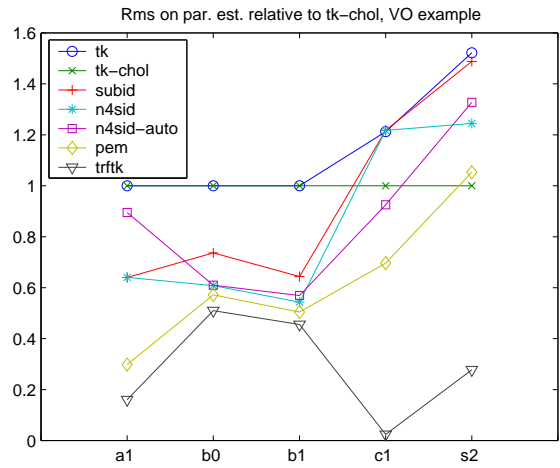


Fig. 1. Rms on parameter estimates relative to tk-chol, VO example

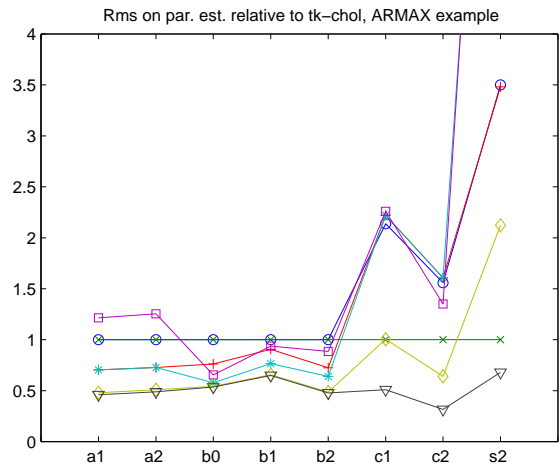


Fig. 2. Rms on parameter estimates relative to tk-chol, ARMAX example. Y-axis truncated.

As uncertainty in the stochastic part is assessed here the focus is on the parameters  $c_1, \dots, c_n$  and  $\sigma^2$ . Based on rms performance for these parameters in the three test cases it is clear that the new tk-chol method is the superior subspace method for the stochastic part but it can not compete with the best prediction error method. Notice also that the automatic choice of  $h$  and  $i$  in n4sid-auto does not improve performance and

System	$a_1$	$a_2$	$a_3$	$a_4$	$b_0$	$b_1$	$b_2$	$b_3$	$b_4$	$c_1$	$c_2$	$c_3$	$c_4$	$\sigma^2$
VO	-0.9490				-2.08950	3.624				-0.9996				6.7050
rms	0.0041				0.25689	0.303				0.2048				1.5246
ARMAX	-1.6464	0.730			0	0.044	0.0398			-1.4026	0.730			0.0030
rms	0.0130	0.011			0.00649	0.010	0.0096			0.0699	0.116			2.8e-4
BJ	-2.8148	3.134	-1.64	0.344	0.10000	-0.140	0.2067	-0.252	0.108	-0.3864	-0.546	0.461	0	0.0001
rms	0.2205	0.458	0.35	0.098	0.00070	0.022	0.0149	0.034	0.031	0.2079	0.076	0.086	0.091	1.67e-5

Table 1. System parameters and rms values for the tk-chol method.

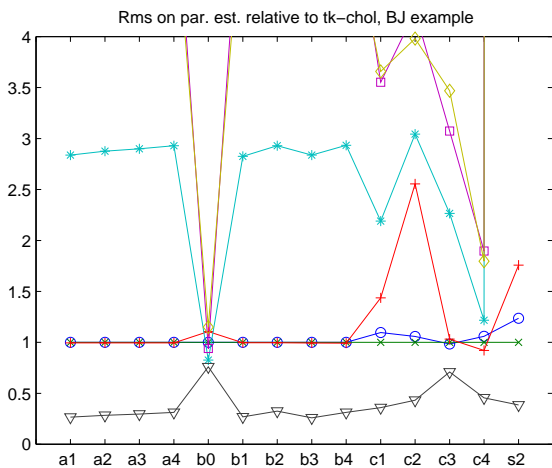


Fig. 3. Rms on parameter estimates relative to tk-chol, BJ example. Y-axis truncated.

that the tk-chol methods is very much better than the other methods in some cases. For the deterministic part, i.e. parameters in  $A(q)$  and  $B(q)$ , the tk-chol method is only the best subspace method for the BJ case. Simulation results not show here reveals that the tk-chol method has the smallest bias for 500 samples but still has asymptotic bias as expected.

## 7. CONCLUSION

In subspace identification the focus has so far been on estimating the deterministic part of the system in question. The result is well performing consistent methods. However, for the stochastic part the existing methods gives large uncertainty partly due to bias. This paper discusses possible improvement for the stochastic part compared to the existing methods. The development of two new methods indicates reduced uncertainty especially for the one using Cholesky factorization. This is verified by simulation for three test cases.

## 8. REFERENCES

- Bauer, D. and M. Jansson (2000). Analysis of the asymptotic properties of the moesp type of subspace algorithms. *Automatica* **36**(4), 497–509.
- Bauer, Dietmar (1998). Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms. PhD thesis. Institute of Econometrics, Operations Research and System Theory, Vienna University of Technology.

Knudsen, Torben (1996). The initialization problem in parameter estimation for general SISO transfer function models. In: *13th World Congress of IFAC* (Janos J. Gertler, Jr. Jose B. Cruz and Michael Peshkin, Eds.). Vol. J. International Federation of Automatic Control. Elsevier. pp. 221–226.

Knudsen, Torben (2001). Consistency analysis of subspace identification methods based on a linear regression approach. *Automatica* **37**(1), 81–89.

Ljung, Lennart (1999). *System Identification, Theory for the User*. Prentice-Hall information and system sciences series. second ed.. Prentice-Hall.

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

van Overschee, Peter and Bart De Moor (1996). *Subspace Identification for Linear Systems*. Kluwer Academic Publishers.