ELSEVIER

A novel subspace identification approach with enforced causal models[☆]

S. Joe Qin[a,*], Weilu Lin[a], Lennart Ljung[b]

[a]*Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, USA*
[b]*Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden*

Abstract

Subspace identification methods (SIMs) for estimating state-space models have been proven to be very useful and numerically efficient. They exist in several variants, but all have one feature in common: as a first step, a collection of high-order ARX models are estimated from vectorized input–output data. In order not to obtain biased estimates, this step must include future outputs. However, all but one of the submodels include non-causal input terms. The coefficients of them will be correctly estimated to zero as more data become available. They still include extra model parameters which give unnecessarily high variance, and also cause bias for closed-loop data. In this paper, a new model formulation is suggested that circumvents the problem. Within the framework, the system matrices (A, B, C, D) and Markov parameters can be estimated separately. It is demonstrated through analysis that the new methods generally give smaller variance in the estimate of the observability matrix and it is supported by simulation studies that this gives lower variance also of the system invariants such as the poles.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Subspace identification; Causal model; Variance analysis

1. Introduction

Subspace identification methods (SIMs) are attractive not only because of their numerical simplicity and stability, but also for their state-space form that is very convenient for optimal estimation, filtering, prediction, and control. Most SIMs fall into the unifying theorem proposed by Van Overschee and De Moor (1995), among which are canonical variate analysis (CVA) (Larimore, 1990, 1992, 2004), N4SID (Van Overschee & De Moor, 1994), subspace fitting (Jansson & Wahlberg, 1996) and MOESP (Verhaegen & Dewilde, 1992). Based on the unifying theorem, all these algorithms can be interpreted as a singular value decomposition of a weighted matrix. The statistical properties such as consistency and efficiency of them have been investigated recently (Bauer, 2003, 2005; Bauer & Ljung, 2002; Chiuso & Picci, 2004; Gustafsson, 2002; Jansson & Wahlberg, 1998; Knudsen, 2001; Larimore, 1996). All these variants are shown to be generically consistent. For some special cases, it has also been shown that CVA gives statistical efficiency and/or gives the lowest variance among available weighting choices. Simulations also seem to indicate that CVA may have better variance properties in overall comparisons, see, e.g. Ljung (2003).

SIMs have many advantages as an alternative to the more traditional prediction error method (PEM) or maximum likelihood (ML) approach and they are very good for delivering initial estimates to PEM. A few drawbacks have been experienced with SIMs

1. The estimation accuracy in general is not as good as the PEM, in terms of the variance of the estimates.
2. The application of SIMs to closed-loop data typically gives biased estimates, even though the data satisfy identifiability conditions for traditional methods such as PEMs.
3. The estimation of B and D may be more problematic than that of A and C, which is reflected in the poor estimation of zeros and steady-state gains (Wang & Qin, 2002).

In this paper, we are concerned with the reasons why subspace identification approaches exhibit these drawbacks and propose new SIMs which use fewer estimated parameters (i.e., more parsimonious) for open-loop applications. First of all, we start with the analysis of existing subspace formulation using the linear regression formulation (Jansson & Wahlberg, 1998; Knudsen, 2001). This means that essentially several ARX models are estimated directly from data with different prediction intervals. From this analysis we reveal that the typical SIM algorithms use extra terms in the model that appear to be non-causal. These terms, although conveniently included for performing subspace projections, are the causes for inflated variance in the estimates and partially responsible for the loss of closed-loop identifiability. Peternell, Scherrer, and Deistler (1996) observe this point as well and use constrained least squares (LS) to improve the estimate. Shi and Macgregor (2001), Jansson (2003), and Larimore (2004) enforce the triangular or causal model structure through pre-estimating the Markov parameters using a high-order ARX model. The proposed algorithms in this paper which extends Qin and Ljung (2003b) do not require a pre-estimation step.

The rest of the paper is organized as follows. In Section 2, we analyze the existing SIMs and point out the non-causal projection. Based on this observation, novel SIM formulations with only causal terms are presented in detail in Section 3. Numerical implementation of proposed algorithms is introduced in Section 4. In Section 5, numerical simulations are given to show the efficiency of the proposed algorithm. Section 6 concludes the paper.

2. Analysis of subspace formulation

2.1. Problem formulation and assumptions

We assume that the system to be identified can be written in an innovation form as

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

$$y_k = Cx_k + Du_k + e_k, \tag{1b}$$

where $y_k \in R^{n_y}$, $x_k \in R^n$, $u_k \in R^{n_u}$, and $e_k \in R^{n_y}$ are the system output, state, input, and innovation, respectively. A, B, C and D are system matrices with appropriate dimensions. K is the Kalman filter gain. To establish statistical consistency of the SIM, we introduce following assumptions:

(A1) The eigenvalues of $A - KC$ are strictly inside the unit circle.

(A2) The system is minimal in the sense that (A, C) is observable and $(A, [B, K])$ is controllable.

(A3) The innovation sequence e_k is a stationary, zero mean, white-noise process with second order moments

$$E(e_i e_j^{\mathrm{T}}) = R\delta_{ij},$$

where δ_{ij} is the Kronecker delta.

(A4) The input u_k and innovation sequence e_j are uncorrelated for $\forall k$ and $\forall j$, i.e., the system operates in open loop.

(A5) The input signal is quasi-stationary (Ljung, 1999b) and is persistently exciting of order $f + p$, where f and p stand for future and past horizons, respectively, to be defined later.

The identification problem is: given a set of input/output measurements, estimate the system matrices (A, B, C, D), Kalman filter gain K up to within a similarity transformation, and the innovation covariance matrix R.

Based on the state-space description in (1), an extended state-space model can be formulated as

$$Y_f = \Gamma_f X_k + H_f U_f + G_f E_f, \tag{2a}$$

$$Y_p = \Gamma_p X_{k-p} + H_p U_p + G_p E_p, \tag{2b}$$

where the subscripts f and p denote future and past horizons, respectively. The extended observability matrix is

$$\Gamma_f = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix} \tag{3}$$

and H_f and G_f are Toeplitz matrices

$$H_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}B & CA^{f-3}B & \cdots & D \end{bmatrix}, \tag{4a}$$

$$G_f = \begin{bmatrix} I & 0 & \cdots & 0 \\ CK & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}K & CA^{f-3}K & \cdots & I \end{bmatrix}. \tag{4b}$$

The input and output data are arranged in the following Hankel form:

$$U_f = \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+N-1} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k+f-1} & u_{k+f} & \cdots & u_{k+f+N-2} \end{bmatrix} \tag{5a}$$

$$\triangleq [u_f(k) \ u_f(k+1) \ \cdots \ u_f(k+N-1)], \tag{5b}$$

$$U_p = \begin{bmatrix} u_{k-p} & u_{k-p+1} & \cdots & u_{k-p+N-1} \\ u_{k-p+1} & u_{k-p+2} & \cdots & u_{k-p+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k-1} & u_k & \cdots & u_{k+N-2} \end{bmatrix} \tag{5c}$$

$$\triangleq [u_p(k-p) \ u_p(k-p+1) \ \cdots \ u_p(k-p+N-1)]. \tag{5d}$$

Similar formulations are made for Y_f, Y_p, E_f, and E_p. The state sequences are defined as

$$X_k = [x_k, x_{k+1}, \ldots, x_{k+N-1}], \tag{6a}$$

$$X_{k-p} = [x_{k-p}, x_{k-p+1}, \ldots, x_{k-p+N-1}]. \tag{6b}$$

Subspace identification consists of estimating the extended observability matrix first and then the model parameters.

2.2. Analysis of conventional SIMs

As the first step, SIMs minimize the following objective function (Van Overschee & De Moor, 1996):

$$[\hat{L}^1 \ \hat{L}^2 \ \hat{L}^3] = \arg\min\{\|Y_f - L^1 Y_p - L^2 U_p - L^3 U_f\|_F^2\}$$

$$= \arg\min\left\{\sum_{j=0}^{N-1}\left\|y_f(k+j)\right.\right.$$

$$\left.\left. - [L^1 \ L^2 \ L^3]\begin{bmatrix} y_p(k-p+j) \\ u_p(k-p+j) \\ u_f(k+j) \end{bmatrix}\right\|^2\right\}, \quad (7)$$

where u_f, u_p, y_f, and y_p are defined in (5b) and (5d) as columns of the corresponding data matrices.

Denoting

$$L^1 = \begin{bmatrix} L^1_{11} & L^1_{12} & \cdots & L^1_{1p} \\ L^1_{21} & L^1_{22} & \cdots & L^1_{2p} \\ \vdots & & \ddots & \\ L^1_{f1} & L^1_{f1} & & L^1_{fp} \end{bmatrix} \triangleq \begin{bmatrix} L^1_1 \\ L^1_2 \\ \vdots \\ L^1_f \end{bmatrix}, \quad (8a)$$

$$L^2 = \begin{bmatrix} L^2_{11} & L^2_{12} & \cdots & L^2_{1p} \\ L^2_{21} & L^2_{22} & \cdots & L^2_{2p} \\ \vdots & & \ddots & \\ L^2_{f1} & L^2_{f1} & & L^2_{fp} \end{bmatrix} \triangleq \begin{bmatrix} L^2_1 \\ L^2_2 \\ \vdots \\ L^2_f \end{bmatrix}, \quad (8b)$$

$$L^3 = \begin{bmatrix} L^3_{11} & L^3_{12} & \cdots & L^3_{1f} \\ L^3_{21} & L^3_{22} & \cdots & L^3_{2f} \\ \vdots & & \ddots & \\ L^3_{f1} & L^3_{f1} & & L^3_{ff} \end{bmatrix} \triangleq \begin{bmatrix} L^3_1 \\ L^3_2 \\ \vdots \\ L^3_f \end{bmatrix}, \quad (8c)$$

the above problem is equivalent to f separate sub-problems:

$$[\hat{L}^1_i \ \hat{L}^2_i \ \hat{L}^3_i]$$

$$= \arg\min\left\{\sum_{j=0}^{N-1}\left\|y_{k+j+i-1}\right.\right.$$

$$\left.\left. - [L^1_i \ L^2_i \ L^3_i]\begin{bmatrix} y_p(k-p+j) \\ u_p(k-p+j) \\ u_f(k+j) \end{bmatrix}\right\|^2\right\} \quad (9)$$

for $i = 1, \ldots, f$, this is to say that f different ARX models are estimated from data. Consider the ith subproblem and spell out the nature of the term $L^3_i u_f(k+j)$. This subproblem

corresponds to the model

$$y_{k+i-1} = [L^1_i \ L^2_i]\begin{bmatrix} y_p(k-p) \\ u_p(k-p) \end{bmatrix} + L^3_i u_f(k) + v_k$$

$$= [L^1_i \ L^2_i]\begin{bmatrix} y_p(k-p) \\ u_p(k-p) \end{bmatrix} + L^3_{i1} u_k + L^3_{i2} u_{k+1}$$

$$+ \cdots + L^3_{ii} u_{k+i-1}$$

$$+ \sum_{j=i+1}^{f} L^3_{ij} u_{k+j-1} + v_k. \quad (10)$$

Note that the summation in (10) represents a non-causal relation from u to y. That is, L^3_{ij} are estimated even though it is known that $L^3_{ij} = 0$ for $j > i$. The matrix L^3 is, in other words, *block lower triangular*. However, this information is not normally taken care of in (7), as pointed out in Shi and Macgregor (2001). While there is no problem from a consistency point of view given proper excitation of the input, known parameters are estimated from data.

Shi (2001) proposes an algorithm known as CVA_{Hf} that removes the impact of future input from the future output using pre-estimated Markov parameters and then performs subspace projections. Shi (2001) further shows that this procedure achieves consistency. Larimore (2004) argues that the CVA_{Hf} was implemented in Adaptx and that it is efficient, but he does not discuss the impact of imperfect pre-estimates. Therefore, we can make the following statements about the typical SIM formulation in general.

1. The model format used in SIM during the projection step is non-causal. This would result in non-causal models in the projection step. Although the non-causal terms are ignored at the step to estimate B, D, all the model parameters estimate have inflated variance due to the fact that extra and unnecessary terms are included in the model.
2. Because of the extra terms that turn out to be 'future' inputs relative to the output, SIMs in general have problems with closed-loop data using direct identification methods. Most SIMs usually project out U_f as follows:

$$Y_f \Pi^{\perp}_{U_f} = \Gamma_f X_k \Pi^{\perp}_{U_f} + G_f E_f \Pi^{\perp}_{U_f}, \quad (11)$$

where $\Pi^{\perp}_{U_f} = I - U_f^T (U_f U_f^T)^{-1} U_f$. Because of the non-causal terms in the model, $(1/N)E_f U_f^T \neq 0$ as $N \to \infty$ for closed-loop data. As a consequence, many SIMs fail to work on closed-loop data, except for a few SIM algorithms that avoid this projection (Chou & Verhaegen, 1997; Wang & Qin, 2002).
3. Because U_f contains extra rows due to the extra terms, the projection in (11) tends to reduce the information content unnecessarily even for open-loop data, leading to inefficient use of the data.
4. These non-causal terms will have negligible coefficients only when the number of data is very large and process is well excited. For a limited number of samples or non-white input signals, SIM algorithms tend to have large estimation errors.

To avoid these problems the SIM model must not include these non-causal terms, Peternell et al. (1996) propose a few methods to exclude these extra terms. Specially, they recommend a two steps procedure: (i) use a conventional (unconstrained) SIM to estimate the deterministic Markov parameters $CA^{i-1}B$; and (ii) form H_f with these Markov parameters to ensure that it is lower triangular and then estimate the extended observability matrix. We propose a parallel and a sequential implementation of a causal subspace identification method (PARSIM) which remove these non-causal terms by enforcing a lower triangular structure in L^3 and hence of H_f at every step of the SIM procedure. By enforcing a lower-triangular structure, we reduce the number of estimated parameters in this stage by $f(f-1)/2$. The parallel PARSIM (PARSIM-P) method involves a bank of LS problems in parallel, while the sequential PARSIM (PARSIM-S) involves a bank of LS problems sequentially. Optimal weighting is derived for the PARSIM algorithms. An optimal estimate of the B, D matrices is given using the Kalman filter structure.

3. Subspace identification avoiding non-causal terms

The key idea in the proposed method is to exclude the non-causal terms of U_f mentioned in Section 2. To accomplish this, we partition the extended state space model row-wise as follows:

$$Y_f = \begin{bmatrix} Y_{f1} \\ Y_{f2} \\ \vdots \\ Y_{ff} \end{bmatrix}, \quad Y_i \triangleq \begin{bmatrix} Y_{f1} \\ Y_{f2} \\ \vdots \\ Y_{fi} \end{bmatrix}, \quad i = 1, 2, \ldots, f, \tag{12}$$

where $Y_{fi} = [y_{k+i-1} \ y_{k+i} \ \cdots \ y_{k+N+i-2}]$. Partition U_f and E_f in a similar way to define U_{fi}, U_i, E_{fi}, and E_i, respectively, for $i = 1, 2, \ldots, f$. Denote further

$$\Gamma_f = \begin{bmatrix} \Gamma_{f1} \\ \Gamma_{f2} \\ \vdots \\ \Gamma_{ff} \end{bmatrix}, \tag{13a}$$

$$H_{fi} \triangleq [CA^{i-2}B \ \cdots \ CB \ D] \tag{13b}$$

$$\triangleq [H_{i-1} \ \cdots \ H_1 \ H_0], \tag{13c}$$

$$G_{fi} \triangleq [CA^{i-2}K \ \cdots \ CK \ I] \tag{13d}$$

$$\triangleq [G_{i-1} \ \cdots \ G_1 \ G_0], \tag{13e}$$

where $\Gamma_{fi} = CA^{i-1}$, and H_i and G_i are the Markov parameters for the deterministic input and innovation sequence, respectively. We have the following equations by partitioning (2a),

$$Y_{fi} = \Gamma_{fi}X_k + H_{fi}U_i + G_{fi}E_i \tag{14}$$

for $i = 1, 2, \ldots, f$. Note that each of the above equations is guaranteed causal.

3.1. PARSIM algorithms

By eliminating e_k in the innovation model through iteration, it is straightforward to derive the following relation (Knudsen, 2001):

$$X_k = L_z Z_p + A_K^p X_{k-p}, \tag{15}$$

where

$$L_z \triangleq [\Delta_p(A_K, K) \ \Delta_p(A_K, B_K)], \tag{16a}$$

$$\Delta_p(A, B) \triangleq [A^{p-1}B \ \cdots \ AB \ B], \tag{16b}$$

$$A_K \triangleq A - KC, \tag{16c}$$

$$B_K \triangleq B - KD, \tag{16d}$$

$$Z_p \triangleq [Y_p^T \ U_p^T]^T. \tag{16e}$$

Substituting this equation into (14), we obtain

$$Y_{fi} = \Gamma_{fi}L_z Z_p + \Gamma_{fi}A_K^p X_{k-p} + H_{fi}U_i + G_{fi}E_i \tag{17}$$

for $i = 1, 2, \ldots, f$. Note the second term on the RHS of (17) tends to zero as p tends to infinity under Assumption A1. Now we have the following parallel PARSIM algorithm to estimate Γ_{fi} and H_{fi}.

Algorithm 1. Parallel PARSIM (PARSIM-P)

1. Perform the following LS estimates, for $i = 1, 2, \ldots, f$,

$$[\hat{\Gamma}_{fi}L_z \ \hat{H}_{fi}] = Y_{fi} \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^\dagger \tag{18}$$

   where $[\cdot]^\dagger$ stands for the Moore–Penrose pseudo-inversion. Stack $\hat{\Gamma}_{fi}L_z$ together to obtain $\hat{\Gamma}_f L_z$ as

$$\begin{bmatrix} \hat{\Gamma}_{f1}L_z \\ \hat{\Gamma}_{f2}L_z \\ \vdots \\ \hat{\Gamma}_{ff}L_z \end{bmatrix} = \hat{\Gamma}_f L_z. \tag{19}$$

2. Perform SVD for the following weighted matrix:

$$W_1(\hat{\Gamma}_f L_z)W_2 = U_n S_n V_n^T + \varepsilon, \tag{20}$$

   where W_1 is non-singular and $L_z W_2$ does not lose rank. U_n, S_n and V_n are associated to the first n largest singular values. The residual term ε stands for the product of the remaining singular vectors and singular values. We choose

$$\hat{\Gamma}_f = W_1^{-1} U_n S_n^{1/2} \tag{21}$$

   from which the estimate of A and C can be obtained (Verhaegen, 1994).

3. The estimate of B and D is discussed in the next section using a Kalman filter formulation.

Note that the proposed parallel PARSIM gives *consistent* estimates for Γ_f and H_{i-1}, $\forall i = 1, 2, \ldots, f$ under the

assumptions stated in Section 2. To rationalize the statement, it is sufficient to show that as $N \to \infty$,

$$[\hat{\Gamma}_{fi}L_z \ \ \hat{H}_{fi}] \to [\Gamma_{fi}L_z \ \ H_{fi}], \tag{22}$$

where $[\hat{\Gamma}_{fi}L_z \ \ \hat{H}_{fi}]$ is calculated according to (18). Assumption A1 implies that the initial state has negligible effect on the estimate with sufficient large p, as shown in (17). From A4 we have $(1/N)E_i Z_p^T \to 0$ and $(1/N)E_i U_i^T \to 0$ as $N \to \infty$. Substituting (17) with $p \to \infty$ into (18) leads to

$$[\hat{\Gamma}_{fi}L_z \ \ \hat{H}_{fi}] = [\Gamma_{fi}L_z \ \ H_{fi}] \begin{bmatrix} Z_p \\ U_i \end{bmatrix} \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^\dagger + G_{fi}E_i \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^\dagger$$

$$= [\Gamma_{fi}L_z \ \ H_{fi}] + G_{fi}\left(\frac{1}{N} E_i \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^T \right)$$

$$\times \left(\frac{1}{N} \begin{bmatrix} Z_p \\ U_i \end{bmatrix} \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^T \right)^{-1}$$

$$\to [\Gamma_{fi}L_z \ \ H_{fi}]$$

as $N \to \infty$. Assumption A5 guarantees that all system modes are sufficiently excited so that the matrix inverse in the above equation exists. It has been shown in Knudsen (2001) that A2 is needed for L_z to have full row rank and Γ_f to have full column rank. Therefore, the SVD step in the PARSIM-P algorithm guarantee that $\hat{\Gamma}_f$ and Γ_f have the same column space asymptotically.

The PARSIM-P algorithm estimates the model parameters in parallel which re-estimate some of the Markov parameters in H_{fi} repeatedly. To avoid this we rewrite (17) by ignoring the A_K^p term

$$Y_{fi} = \Gamma_{fi}L_z Z_p + H_{i-1}U_{f1} + H_{f(i-1)}[U_{f2}^T \ \ \cdots \ \ U_{fi}^T]^T$$
$$+ G_{fi}E_i,$$

where H_{i-1} is defined in (13c). If we perform the above projections sequentially for $i = 1, 2, \ldots, f$, $H_{f(i-1)}$ is estimated in the $(i-1)$th step. Γ_{fi} and H_{i-1} are the only unknown at the ith step.

Algorithm 2. Sequential PARSIM (PARSIM-S)

1. Perform the following LS for $i = 1$,

$$[\hat{\Gamma}_{f1}L_z \ \ \hat{H}_{f1}] = Y_{f1} \begin{bmatrix} Z_p \\ U_{f1} \end{bmatrix}^\dagger. \tag{23}$$

2. Perform the following causal projection for $i = 2, \ldots, f$

$$[\hat{\Gamma}_{fi} \ \ \hat{H}_{i-1}] = (Y_{fi} - \hat{H}_{f(i-1)})[U_{f2}^T \ \ \cdots \ \ U_{fi}^T]^T$$
$$\times \begin{bmatrix} Z_p \\ U_{f1} \end{bmatrix}^\dagger. \tag{24}$$

Stack $\hat{\Gamma}_{fi}L_z$ together as Eq. (19).
3. Same as Step 2 in Algorithm 1.

The sequential PARSIM gives *consistent* estimates for Γ_f and H_{i-1}, $\forall i = 1, 2, \ldots, f$ under the assumptions stated in

Section 2. The proof is similar to that of PARSIM-P, therefore, we omit it in the paper.

Remark 1. For finite past horizon p the algorithm is biased, but the bias decays to zero exponentially with p. If p is too large in practice, however, large variance is expected for the estimates. Therefore, it is necessary in practice to use a finite p for the best trade-off. Cross-validation can be used to select an optimal p.

Remark 2. The parallel PARSIM requires that no correlation exists between future u_k and past e_k to be consistent, which is only valid under open loop condition, therefore, the PARSIM-P algorithms are biased for direct closed-loop identification. To make it applicable to closed-loop data, an innovation estimation approach is proposed in Qin and Ljung (2003a).

Remark 3. The Markov parameters, H_{i-1}, $\forall i = 1, 2, \ldots, f$, can be estimated directly from the SIMs without the knowledge of system matrices, (A, B, C, D). Meanwhile, the low triangular structure of the Toeplitz matrix, H_f, is conserved.

3.2. Improved variance of PARSIM algorithms

After presenting the PARSIM algorithms, we analyze the variance of the PARSIM estimates relative to that of conventional SIM algorithms. For conventional SIMs the asymptotic variance of the model estimates is derived in Bauer and Jansson (2000), Bauer and Ljung (2002), Chiuso and Picci (2004). These analyses provide insight into what contribute to the variance of the estimates.

In this subsection we provide a covariance equality for PARSIM estimates by interpreting the sub-space projections in the generalized least squares (GLS) framework (Mardia, Kent, & Bibby, 1979). For the ith block row we explained that conventional SIMs use model (10) but the process is actually (17). By comparing (10) with (17) when p is large we have,

$$[L_i^1 \ \ L_i^2] = \Gamma_{fi}L_z,$$

$$L_i^3 = [H_{fi} \ \vdots \ 0 \ \ \cdots \ \ 0],$$

$$v_k = \sum_{j=1}^{i} G_{i-j}e_{k+j-1}.$$

Note that v_k is auto-correlated, therefore, the SIM projections do not fit into the ML framework. Denoting

$$cov[v_k] = \Sigma_v,$$

$$V_{fi} = G_{fi}E_i$$

and

$$\mathcal{V}_{i,N} = vec(V_{fi}^T),$$

where $vec()$ of a matrix forms a long column vector by stacking the columns of that matrix, we have

$$cov(\mathcal{V}_{i,N}) = \Sigma_v \otimes I_N.$$

With this notation we can convert the PARSIM equation (17) into

$$\mathscr{Y}_{i,N} = \mathscr{X}_{1,N}\theta_1 + \mathscr{V}_{i,N}, \tag{25}$$

where

$$\mathscr{Y}_{i,N} = vec(Y_{fi}^{\mathrm{T}}),$$

$$\mathscr{X}_{1,N} = I \otimes [Z_p^{\mathrm{T}} \ U_i^{\mathrm{T}}],$$

$$\theta_1 = vec([\Gamma_{fi}L_z \ H_{fi}]^{\mathrm{T}}).$$

Similarly, the conventional SIM equation for the ith block row (10) can be converted to

$$\mathscr{Y}_{i,N} = \mathscr{X}_{1,N}\theta_1 + \mathscr{X}_{2,N}\theta_2 + \mathscr{V}_{i,N}, \tag{26}$$

where

$$\mathscr{X}_{2,N} = I \otimes [U_{f(i+1)}^{\mathrm{T}} \ U_{f(i+2)}^{\mathrm{T}} \ \cdots \ U_{ff}^{\mathrm{T}}]$$

is the matrix of non-causal input data and

$$\theta_2 = vec([L_{i(i+1)}^3 \ L_{i(i+2)}^3 \ \cdots \ L_{if}^3]^{\mathrm{T}})$$

is the vector of extra parameters in conventional SIMs. Now we state that the LS solutions (17) for PARSIMs and (10) for conventional SIMs are identical to the GLS solution to (25) and (26), respectively (Mardia et al., 1979). The estimates from both conventional SIMs and PARSIMs are consistent, which is not of concern here. The question is whether PARSIM estimates have smaller variance than conventional SIMs regardless of the data length N.

From Mardia et al. (1979) we know that the GLS interpretation of PARSIM estimates leads to

$$cov(\hat{\theta}_{1,N}) = (\mathscr{X}_{1,N}^{\mathrm{T}}(\Sigma_v \otimes I_N)^{-1}\mathscr{X}_{1,N})^{-1},$$

where $\hat{\theta}_{1,N}$ is the PARSIM estimate for θ_1 and

$$cov\left(\begin{bmatrix}\hat{\theta}_{1,N}' \\ \hat{\theta}_{2,N}'\end{bmatrix}\right)$$
$$= ([\mathscr{X}_{1,N} \ \mathscr{X}_{2,N}]^{\mathrm{T}}(\Sigma_v \otimes I_N)^{-1}[\mathscr{X}_{1,N} \ \mathscr{X}_{2,N}])^{-1},$$

where $\hat{\theta}_{1,N}'$ is the conventional SIM estimate for θ_1.

To simplify the notation, we denote

$$S_{ij,N} = \mathscr{X}_{i,N}(\Sigma_v \otimes I_N)^{-1}\mathscr{X}_{j,N} \tag{27}$$

for $i, j = 1, 2$. Then the covariance expressions become

$$cov(\hat{\theta}_{1,N}) = S_{11,N}^{-1},$$

$$cov\left(\begin{bmatrix}\hat{\theta}_{1,N}' \\ \hat{\theta}_{2,N}'\end{bmatrix}\right) = \begin{bmatrix}S_{11,N} & S_{12,N} \\ S_{12,N}^{\mathrm{T}} & S_{22,N}\end{bmatrix}^{-1} = \begin{bmatrix}\Phi_{11,N} & \Phi_{12,N} \\ \Phi_{12,N}^{\mathrm{T}} & \Phi_{22,N}\end{bmatrix}$$

from which it is easy to show that

$$cov(\hat{\theta}_{1,N}') = \Phi_{11,N} = S_{11,N}^{-1} + S_{11,N}^{-1}S_{12,N}\Phi_{22,N}S_{12,N}^{\mathrm{T}}S_{11,N}^{-1}.$$

Therefore,

$$cov(\hat{\theta}_{1,N}') - cov(\hat{\theta}_{1,N}) = S_{11,N}^{-1}S_{12,N}\Phi_{22,N}S_{12,N}^{\mathrm{T}}S_{11,N}^{-1}. \tag{28}$$

Noticing that $\Phi_{22,N}$ is strictly positive definite due to the inverse of the covariance matrix, we have

$$cov(\hat{\theta}_{1,N}') \geqslant cov(\hat{\theta}_{1,N}) \tag{29}$$

regardless of N and the equality holds only if $S_{12,N} = 0$. It is noted further that $S_{11,N}^{-1}S_{12,N}$ in (28) is the regression coefficient matrix of $\mathscr{X}_{2,N}$ on $\mathscr{X}_{1,N}$, which is not zero for colored inputs. We can only compare the variance of $\hat{\theta}_1$ rigorously as shown above. For $\hat{\Gamma}_f$ estimate from $\hat{\theta}_1$ we can only say that PARSIM estimate likely leads to a better estimate of the true observability subspace, but we cannot compare the variance since it depends on the basis. Similarly we cannot compare the variance of the system matrices such as C and A.

With this we conclude that the PARSIM estimates of the Markov parameters generally have a smaller variance than those of the conventional SIM estimates for any N. The estimates of the A and C matrices are unique functions (in the chosen basis) of $\hat{\theta}_{1,N}$. It thus follows from (29) and the Gauss approximations formula that the asymptotic covariances of A and C are also as good or better with the PARSIM approach. The Monte-Carlo study in Section 5.2 provides strong indications that this indeed is the case.

3.3. Determination of observability matrix with optimal weighting

In the conventional SIM formulation under open-loop conditions,

$$E_f \Pi_{U_f}^{\perp} \to E_f \quad \text{as } N \to \infty \tag{30}$$

since E_f is uncorrelated with U_f. Therefore, for large N (11) becomes

$$Y_f \Pi_{U_f}^{\perp} \approx \Gamma_f X_k \Pi_{U_f}^{\perp} + G_f E_f. \tag{31}$$

Post-multiplying Z_p^{T} to (31) eliminate the noise term for large N,

$$Y_f \Pi_{U_f}^{\perp} Z_p^{\mathrm{T}} \approx \Gamma_f X_k \Pi_{U_f}^{\perp} Z_p^{\mathrm{T}}. \tag{32}$$

Van Overschee and De Moor (1995) show that all SIM methods perform SVD on the following weighted matrix:

$$W_{\mathrm{r}} Y_f \Pi_{U_f}^{\perp} Z_p^{\mathrm{T}} W_{\mathrm{c}} = W_{\mathrm{r}} \Gamma_f X_k \Pi_{U_f}^{\perp} Z_p^{\mathrm{T}} W_{\mathrm{c}}, \tag{33}$$

where W_{r} and W_{c} are the row and column weighting matrices, respectively. In CVA $W_{\mathrm{r}} = (Y_f \Pi_{U_f}^{\perp} Y_f^{\mathrm{T}})^{-1/2}$ which basically normalizes the output variables. Gustafsson (2002) shows that an approximately optimal weighting for W_{c} is

$$W_{\mathrm{c}} = (Z_p Z_p^{\mathrm{T}} - Z_p U_f^{\mathrm{T}}(U_f U_f^{\mathrm{T}})U_f Z_p^{\mathrm{T}})^{-1/2}$$
$$= (Z_p \Pi_{U_f}^{\perp} Z_p^{\mathrm{T}})^{-1/2}, \tag{34}$$

which is used in CVA and MOESP. Substituting (34) into (33), and replacing X_k with $L_z Z_p$ as instrumental variables,

we obtain,

$$W_r Y_f \Pi_{U_f}^{\perp} Z_p^T W_c = W_r \Gamma_f L_z Z_p \Pi_{U_f}^{\perp} Z_p^T (Z_p \Pi_{U_f}^T Z_p^T)^{-1/2}$$
$$= W_r \Gamma_f L_z (Z_p \Pi_{U_f}^{\perp} Z_p^T)^{1/2}. \tag{35}$$

Comparing (35) with (20), the equivalent weighings for the PARSIMs algorithm are

$$W_1 = W_r, \tag{36}$$

$$W_2 = (Z_p \Pi_{U_f}^{\perp} Z_p^T)^{1/2}. \tag{37}$$

Gustafsson and Rao (2002) show that the row-weighting W_1 has no influence on the asymptotic accuracy of the estimated observability matrix. Our simulation experience shows that W_1 has negligible influence on the accuracy of the estimated system matrices as well. Therefore, we suggest to use $W_1 = I$ in the PARSIM algorithms.

4. Numerical implementation of PARSIMs

Since the projections in the PARSIM algorithms bear similarity to the standard SIMs such as MOESP, it is straightforward to implement these parallel or sequential projections using QR decomposition (Qin & Ljung, 2003b). In this section, a new approach to calculate the B, D matrices is derived by pre-whitening the equation error of the general state-space model.

4.1. QR implementation for K

Once $\hat{\Gamma}_f$ is known, the Kalman filter gain K can be estimated (Di Ruscio, 1996). With a large p, substituting (15) into (2) leads to

$$Y_f = \Gamma_f L_z Z_p + H_f U_f + G_f E_f. \tag{38}$$

Therefore,

$$Y_f \Pi_{\substack{Z_p \\ U_f}}^{\perp} = G_f E_f \Pi_{\substack{Z_p \\ U_f}}^{\perp} = G_f E_f \tag{39}$$

since E_f is not correlated with Z_p and U_f in open loop. Performing QR decomposition,

$$\begin{bmatrix} Z_p \\ U_f \\ Y_f \end{bmatrix} = \begin{bmatrix} R_{11} & & \\ R_{21} & R_{22} & \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \tag{40}$$

then

$$R_{33} Q_3 = G_f E_f. \tag{41}$$

Denoting $e_k = F e_k^*$ such that $cov(e_k^*) = I$, from Assumption A3 we have $F F^T = R$. Using this notation we have

$$G_f E_f = G_f^* E_f^*, \tag{42}$$

where

$$G_f^* = \begin{bmatrix} F & 0 & \cdots & 0 \\ CKF & F & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}KF & CA^{f-3}KF & \cdots & F \end{bmatrix}$$
$$\in \Re^{n_y f \times n_y f}. \tag{43}$$

From Eqs. (41) and (42) and using the fact that Q_3 is an orthonormal matrix, we choose

$$\hat{E}_f^* = Q_3, \tag{44a}$$

$$\hat{G}_f^* = R_{33}. \tag{44b}$$

Therefore,

$$\hat{F} = R_{33}(1 : n_y, 1 : n_y) \tag{45}$$

and K can be calculated from G_f^* using Γ_f.

4.2. Determination of B, D

With A and C estimates, Section 10.6 in Ljung (1999b) gives an effective approach to estimate B and D with an output error formulation. Note that there is a choice whether or not to pre-whiten the residuals, as discussed, e.g., in Ljung (1999a). This choice also corresponds to whether 'focus' is set to 'simulation' or 'prediction' (default) in the N4SID function of the System Identification Toolbox. Here, we give a modified approach to estimating B, D and the initial state optimally using A, C, K and F for the general innovation form. Since the initial state is estimated this step does not introduce a bias for finite p.

From the innovation form of the system we have

$$x_{k+1} = A_K x_k + B_K u_k + K y_k, \tag{46}$$

where A_k and B_k are defined in (16). The process output can be represented as

$$y_k = C(qI - A_K)^{-1}x_0 + [C(qI - A_K)^{-1}B_K + D]u_k$$
$$+ C(qI - A_K)^{-1}K y_k + e_k \tag{47}$$

or

$$[I - C(qI - A_K)^{-1}K]y_k$$
$$= C(qI - A_K)^{-1}x_0 + [C(qI - A_K)^{-1}B_K + D]u_k + e_k \tag{48}$$

using $e_k = F e_k^*$ where e_k^* has an identity covariance matrix, and defining

$$\tilde{y}_k = F^{-1}[I - C(qI - A_K)^{-1}K]y_k, \tag{49a}$$

$$G(q) = F^{-1}C(qI - A_K)^{-1}, \tag{49b}$$

$$D^* = F^{-1}D \tag{49c}$$

we obtain,

$$\tilde{y}_k = G(q)B_K u_k + D^* u_k + G(q)x_0 \delta_k + e_k^*$$
$$= G(q) \otimes u_k^T vec(B_K) + I_{n_y} \otimes u_k^T vec(D^*)$$
$$+ G(q)x_0 \delta_k + e_k^*, \tag{50}$$

where $vec(B_K)$ and $vec(D^*)$ are vectorized B_K and D^* matrices along the rows. δ_k is the Kronecker delta function. Now $vec(B_K)$, $vec(D^*)$ and x_0 can be estimated using LS from the above equation. The B, D matrices can be backed out as

$$\hat{D} = F \hat{D}^*, \tag{51a}$$

$$\hat{B} = \hat{B}_K + K \hat{D}. \tag{51b}$$

5. Simulation and industrial case studies

In this section, the results of two simulation cases and an industrial case are reported to demonstrate the efficiency of proposed PARSIMs with comparison to N4SID in the System Identification Toolbox (Version 6.0) of Matlab. The first simulation is a second order single input and single output (SISO) counter example from Jansson and Wahlberg (1998). The second is a Monte-Carlo simulation study over randomly chosen fourth order systems with two inputs and two outputs. The industrial case study is a 3×3 four-stage evaporator from Van Overschee and De Moor (1996).

5.1. Simulation example 1

The counter-example proposed in Jansson and Wahlberg (1998) is used here to test the effectiveness of the proposed parallel PARSIM methods

$$x_{k+1} = \begin{bmatrix} 2\gamma & -\gamma^2 \\ 1 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ -2 \end{bmatrix} u_k + \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} e_k, \tag{52a}$$

$$y_k = [2 \quad -1] x_k + e_k, \tag{52b}$$

where the variance of the noise process $var(e_k) = 217.1$, $\gamma = 0.9184$, $k_1 = -0.21$ and $k_2 = -0.559$ are used here. The system input is a high-pass filter with unit white Gaussian noise as input.

$$u_k = (1 - \gamma q^{-1})^2 (1 + \gamma q^{-1})^2 \varepsilon_k.$$

For comparison we use the N4SID routine in Matlab, which actually implemented the CVA weighting, as the standard SIM algorithm. PEM implemented as the ARMAX routine in Matlab's System Identification Toolbox is used as a benchmark. The performance of the methods is investigated with two indices, the standard deviation of the pole estimation errors and that of the zero estimation errors,

$$P(N) = \frac{1}{M} \sum_{k=1}^{M} \| \hat{P}_N^k - P_0 \|_2, \tag{53a}$$

$$Z(N) = \frac{1}{M} \sum_{k=1}^{M} \| \hat{Z}_N^k - Z_0 \|_2, \tag{53b}$$

where $M = 200$ is the number of independent simulations. $\hat{P}_N^k$ and $\hat{Z}_N^k$ are the estimated poles and zeros with N samples at kth simulation, respectively. P_0 and Z_0 are the true poles and zeros of the system, respectively. We choose $p = 7$, $f = 5$ for PARSIMs. The results of the simulations are shown in Figs. 1 and 2, which show the asymptotical performance of different algorithms. The results show that the PARSIMs outperform N4SID for both pole and zero estimation, and the zero estimates of PARSIMs are very close to those of PEM.

5.2. Simulation example 2

To study the potential benefits of the causal parameterization in PARSIMs, we perform a Monte-Carlo study over randomly
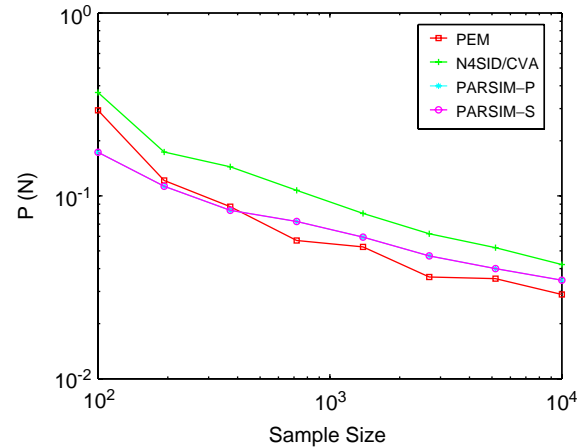

Fig. 1. Asymptotic pole estimation results of the SISO counter example.
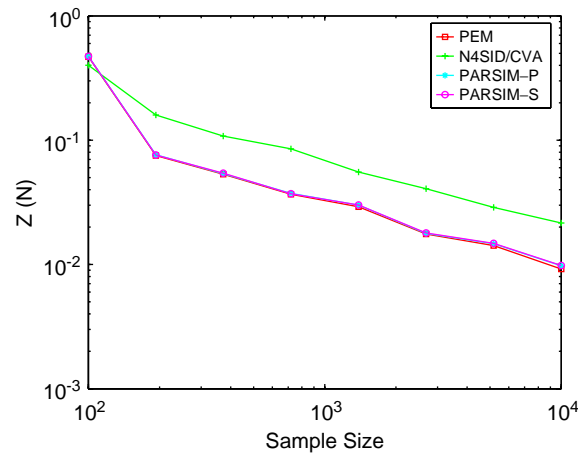

Fig. 2. Asymptotic zero estimation results of the SISO counter example.

chosen fourth order systems with two inputs and two outputs, estimated with the different methods. Since the motivation for the causal parameterization is to provide a better estimate of the observability matrix, we concentrate on the estimates of the A-matrix, viz. its eigenvalues. The input was chosen as a random binary signal with power up to 0.1 of the Nyquist frequency, and normal white noise with 0.1 times the unit covariance matrix was added to the output.

The system and the input/output data were generated by Matlab as follows:

```
m0 = idss(drss(4,2,2));
m0.d = zeros(2,2);
m0.b = 5*randn(4,2);
u = idinput([400,2],'rbs',[0 0.1]);
y = sim(m0,u) + 0.1*randn(400,2);
```

For each data set a model was estimated using Matlab's standard N4SID/CVA routine as well as using PARSIM-S and
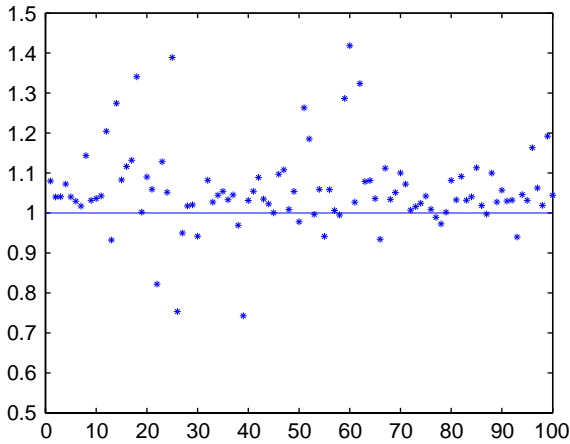
Fig. 3. The value r defined by (54) for the 100 randomly chosen systems as defined in the text. The number of realizations for each system, M, was 25. (The value for system 31 is 4.46 and out of range.) PARSIM-P is better than N4SID/CVA in 84 of the 100 cases. The average excess of standard deviation for N4SID/CVA is 9.1%.

PARSIM-P. The future horizon (f) was chosen as 20 and the past horizon (p) was chosen as 10 in all cases.

For each method the standard deviations (absolute values) of each of the four eigenvalues were estimated over the M realizations in the usual way

$$\hat{\sigma}_i^{\text{CVA}}, \quad \hat{\sigma}_i^{\text{PARSIM-P}}, \quad i = 1, 2, 3, 4.$$

As PARSIM-S and PARSIM-P gave nearly the same results, we only compare the performance of N4SID/CVA and PARSIM-P in the paper. The mean of the ratios of the accuracy of the methods was computed

$$r = \frac{1}{4} \sum_{i=1}^{4} \frac{\hat{\sigma}_i^{\text{CVA}}}{\hat{\sigma}_i^{\text{PARSIM-P}}} \tag{54}$$

as a measure of the relative accuracy of the two methods in estimating the poles/eigenvalues. A plot of r over the 100 different systems with $M = 25$ is given in Fig. 3. It is seen that the enforcement of the causal model when estimating the observability matrix gives a noticeable improvement in the standard deviation of the eigenvalue estimates.

5.3. Industrial case study

In this subsection, the experimental data from a four-stage evaporator are analyzed (Van Overschee & De Moor, 1996). The three inputs are feed flow, vapor flow to the first evaporator stage and cooling water flow. The three outputs are the dry matter content, the flow and temperature of the outcoming product. The time series plot of the data indicates that the inputs are PRBS. There are 6305 experimental data points, we use the first 3152 points for estimation and the rest of them for validation. We choose $p = 30$, $f = 20$ for PARSIMs and N4SID. By using the

Table 1
The model fit as measured by R^2 in (55) of identified models for simulation and prediction of validation data for the evaporator. (R^2 less than zero is indicated by '–'.)

	1 step ahead	20 steps ahead	100 steps ahead	Simulation	
y_1	74.79	60.16	49.64	44.35	N4SID
	74.65	61.20	54.14	51.24	PARSIM-P
	74.52	60.75	53.87	48.16	PARSIM-S
y_2	62.12	60.27	60.12	58.15	N4SID
	61.62	60.07	60.28	59.97	PARSIM-P
	61.43	60.43	60.01	59.55	PARSIM-S
y_3	84.50	57.27	14.27	—	N4SID
	84.47	60.60	42.06	30.35	PARSIM-P
	84.47	60.64	34.49	15.35	PARSIM-S

Akaike information criterion (AIC) and examining the singular values an 11th order system is chosen.

The coefficient of determination

$$R^2 = \left(1 - \frac{\Sigma_i [y_k(i) - \hat{y}_k(i)]^2}{\Sigma_i [y_k(i) - \bar{y}_k]^2} \right) \times 100 \tag{55}$$

of validation data is used as the metric for comparing different SIMs, where y_k, $\hat{y}_k$ and $\bar{y}_k$ are the measured output, simulated or predicted model output and the mean of the output for the kth output variables, respectively. The result of simulation and various horizons of prediction for different SIMs is shown in Table 1.

From the result, we can see that all methods work well for one-step ahead predictions. As the prediction horizon increases, the prediction accuracy decreases, as expected. In general, the PARSIM algorithms outperform the N4SID on long-term predictions. For simulation error N4SID failed on y_3. While the N4SID results are almost the same as the PARSIM results for 1 and 20 steps ahead predictions, the PARSIMs produce better results for 100 steps ahead prediction and simulation.

6. Conclusions

In this paper, a novel sub-space identification approach is proposed to enforce the casuality of high-order ARX models. The key idea is to avoid the estimation of parameters that are known to be zero. This means that a lower triangular structure of an estimated matrix must be enforced which leads to somewhat more complicated calculations. Also other authors have noted the potential problems that arise from these non-causal elements. Ljung and McKelvey (1996) have noted that the problems with closed-loop data have their roots in these non-causal terms. They suggest to use explicitly computed k-step ahead predictions from a single causal ARX-model. This is very different from the algorithm suggested in this paper, and apparently it does not make the best use of the observed data. The new algorithms, which fall into the subspace fitting framework, are shown to be consistent under mild assumptions and applicable to a general state-space model structure.

We have shown that the variance of the observability matrix estimates is in general smaller if the non-causal terms are omitted. It is difficult to make further comparison about the variance of the system matrices because they depend on the basis. Simulation tests are conducted to compare the variance of the eigenvalues of the A matrix. We have indeed seen improved behavior in the reported tests. The simulation studies indicate that the proposed algorithms are superior to SIMs with CVA weighting, which are considered optimal.

Acknowledgments

References

Bauer, D. (2003). Subspace algorithms. In *Proceedings of the 13th IFAC SYSID symposium* (pp. 1030–1041). Rotterdam, Netherlands, August.

Bauer, D. (2005). Asymptotic properties of subspace estimators. *Automatica*, *41*, 359–376.

Bauer, D., & Jansson, M. (2000). Analysis of the asymptotic properties of the MOESP type of subspace algorithm. *Automatica*, *36*, 497–509.

Bauer, D., & Ljung, L. (2002). Some facts about the choice of the weighting matrices in LARIMORE type of subspace algorithms. *Automatica*, *38*, 763–773.

Chiuso, A., & Picci, G. (2004). The asymptotic variance of subspace estimates. *Journal of Econometrics*, *118*, 257–291.

Chou, C. T., & Verhaegen, M. (1997). Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, *33*, 1857–1869.

Di Ruscio, D. (1996). Combined deterministic and stochastic system identification and realization: DSR—a subspace approach based on observations. *Modeling, Identification and Control*, *17*, 193–230.

Gustafsson, T. (2002). Subspace-based system identification: Weighting and pre-filtering of instruments. *Automatica*, *38*, 433–443.

Gustafsson, T., & Rao, B. D. (2002). Statistical analysis of subspace-based estimation of reduced-rank linear regression. *IEEE Transactions of Signal Processing*, *50*, 151–159.

Jansson, M. (2003). Subspace identification and ARX modelling. In *Proceedings of the 13th IFAC SYSID symposium*, Rotterdam, Netherlands, August.

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

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

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

Larimore, W. E. (1990). Canonical variate analysis in identification, filtering and adaptive control. In *IEEE conference on decision and control* (pp. 596–604), December.

Larimore, W. E. (1992). *ADAPTx Automated System Identification Software Users Manual*. Adaptics, Inc.

Larimore, W. E. (1996). Statistical optimality and canonical variate analysis system identification. *Signal Processing*, *52*, 131–144.

Larimore, W. E. (2004). Large sample efficiency for ADAPTx subspace system identification with unknown feedback. In *Proceedings of the 7th international symposium on DYCOPS*, Boston, MA, July.

Ljung, L. (1999a). Estimation focus in system identification: Prefiltering, noise models, and prediction. In *IEEE conference on decision and control* (pp. 2810–2815), December.

Ljung, L. (1999b). *System identification—theory for the user*. 2nd ed., Englewood Cliffs, NJ: Prentice Hall PTR.

Ljung, L. (2003). Aspects and experiences of user choices in subspace identification methods. In *Proceedings of the 13th IFAC SYSID symposium* (pp. 1802–1807). Rotterdam, Netherlands, August.

Ljung, L., & McKelvey, T. (1996). Subspace identification from closed loop data. *Signal Processing*, *52*, 209–215.

Mardia, K., Kent, J., & Bibby, J. (1979). *Multivariate analysis*. New York: Academic Press.

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

Qin, S., & Ljung, L. (2003a). Closed-loop subspace identification with innovation estimation. In *Proceedings of the 13th IFAC SYSID symposium* (pp. 887–892). Rotterdam, Netherlands, August.

Qin, S., & Ljung, L. (2003b). Parallel QR implementation of subspace identification with parsimonious models. In *Proceedings of the 13th IFAC SYSID symposium* (pp. 1631–1636). Rotterdam, Netherlands, August.

Shi, R. (2001). *Subspace identification methods for dynamic process modeling*. Ph.D. thesis, Department of Chemical Engineering, McMaster University.

Shi, R., & Macgregor, J. F. (2001). A framework for subspace identification methods. In *Proceedings of the American control conference* (pp. 3678–3683). Arlington, VA, June.

Van Overschee, P., & De Moor, B. (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, *30*, 75–93.

Van Overschee, P., & De Moor, B. (1995). A unifying theorem for three subspace system identification algorithms. *Automatica*, *31*, 1853–1864.

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

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

Verhaegen, M., & Dewilde, P. (1992). Subspace model identification. Part I: The output-error state-space model identification class of algorithms. *International Journal of Control*, *56*, 1187–1210.

Wang, J., & Qin, S. J. (2002). A new subspace identification approach based on principal component analysis. *Journal of Process Control*, *12*, 841–855.

Dr. S. Joe Qin holds Paul D. and Betty Robertson Meek and American Petrofina Foundation Centennial Professorship in Chemical Engineering at University of Texas at Austin. He obtained his B.S. and M.S. degrees in Automatic Control from Tsinghua University in Beijing, China, in 1984 and 1987, respectively. He received his Ph.D. degree in Chemical Engineering from University of Maryland in 1992. He worked as a Principal Engineer at Fisher-Rosemount from 1992 to 1995 and joined University of Texas as a professor in 1995. His research interests include system identification, process monitoring and fault diagnosis, model predictive control, run-to-run control, semiconductor process control, and control performance monitoring. He is a recipient of the NSF CAREER Award, DuPont Young Professor Award, NSF-China Outstanding Young Investigator Award, and is currently an Editor for *Control Engineering Practice* and a Member of the Editorial Board for *Journal of Chemometrics*.

Weilu Lin received his Ph.D. degree from the Department of Chemical Engineering at University of Texas at Austin in 2005. He received the Bachelor of Engineering in Chemical Engineering with honor in 1994 from Xiamen University. In 1997, he received the Master of Engineering in Chemical Engineering from Nanjing University of Chemical Technology. He is currently a process control engineer at Weyerhaeuser Co. His research interest includes system identification, process monitoring, and process control.

Lennart Ljung received his Ph.D. in Automatic Control from Lund Institute of Technology in 1974. Since 1976 he is Professor of the chair of Automatic Control in Linkoping, Sweden, and is currently Director of the Competence Center "Information Systems for Industrial Control and Supervision" (ISIS). He has held visiting positions at Stanford and MIT and has written several books on System Identification and Estimation. He is an IEEE Fellow and an IFAC Advisor as well as a member of the Royal Swedish Academy of Sciences (KVA), a member of the Royal Swedish Academy of Engineering Sciences (IVA), an Honorary Member of the Hungarian Academy of Engineering and a Foreign Associate of the US National Academy of Engineering (NAE). He has received honorary doctorates from the Baltic State Technical University in St. Petersburg, from Uppsala University, Sweden, from the Technical University of Troyes, France, and from the Catholic University of Leuven, Belgium. In 2002, he received the Quazza Medal from IFAC and in 2003 he received the Hendryk W. Bode Lecture Prize from the IEEE Control Systems Society.