Technical report from Automatic Control at Linköpings universitet

Handling Certain Structure Information in Subspace Identification

Christian Lyzell, Martin Enqvist, Lennart Ljung

Division of Automatic Control

E-mail: lyzell@isy.liu.se, maren@isy.liu.se, ljung@isy.liu.se

3rd April 2009

Report no.: LiTH-ISY-R-2890

Accepted for publication in 15th IFAC Symposium on System Identification, Saint-Malo, France

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Technical reports from the Automatic Control group in Linköping are available from http://www.control.isy.liu.se/publications.

Abstract

The prediction-error approach to parameter estimation of linear models often involves solving a non-convex optimization problem. In some cases, it is therefore difficult to guarantee that the global optimum will be found. A common way to handle this problem is to find an initial estimate, hopefully lying in the region of attraction of the global optimum, using some other method. The prediction-error estimate can then be obtained by a local search starting at the initial estimate. In this paper, a new approach for finding an initial estimate of certain linear models utilizing structure and the subspace method is presented. The polynomial models are first written on the observer canonical state-space form, where the specific structure is later utilized, rendering least-squares estimation problems with linear equality constraints.

Keywords: System identification, Subspace method, Black-box models.

Handling Certain Structure Information in Subspace Identification

C. Lyzell * **M. Enqvist** * **L. Ljung** *

* *Division of Automatic Control, Department of Electrical Engineering, Linköpings Universitet, SE – 581 83, Linköping, Sweden. (E-mail: {lyzell,maren,ljung}@isy.liu.se)*

Abstract: The prediction-error approach to parameter estimation of linear models often involves solving a non-convex optimization problem. In some cases, it is therefore difficult to guarantee that the global optimum will be found. A common way to handle this problem is to find an initial estimate, hopefully lying in the region of attraction of the global optimum, using some other method. The prediction-error estimate can then be obtained by a local search starting at the initial estimate. In this paper, a new approach for finding an initial estimate of certain linear models utilizing structure and the subspace method is presented. The polynomial models are first written on the observer canonical state-space form, where the specific structure is later utilized, rendering least-squares estimation problems with linear equality constraints.

Keywords: System identification, Subspace method, Black-box models.

1. INTRODUCTION

The estimation of linear polynomial models

$$A_p(q)y(t) = \frac{B_p(q)}{F_p(q)}u(t) + \frac{C_p(q)}{D_p(q)}e(t), \tag{1}$$

given a sequence of input and output pairs

$$Z^N \triangleq \big(u(t), y(t)\big)_{t=1}^N, \tag{2}$$

is a classical problem in system identification. Several methods exist, where the *prediction-error method* (PEM) or the *instrumental variable method* (IV) are common choices, see e.g. Ljung (1999); Söderström and Stoica (1989). In the last two decades, the *subspace system identification* methods (SID) have been developed and become an important set of tools for estimating state-space models

$$x(t+1) = A(\theta)x(t) + B(\theta)u(t) + w(t), \tag{3a}$$
$$y(t) = C(\theta)x(t) + D(\theta)u(t) + v(t), \tag{3b}$$

see e.g. Van Overschee and De Moor (1996b); Verhaegen and Verdult (2007); Ljung (1999).

In this paper, we are only going to consider a special case of (1), namely the ARMAX model structure where

$$F_p(q) = D_p(q) = 1.$$

In Bauer (2005), a theoretical analysis of the possibility of using SID methods to estimate ARMAX models is given. The paper concerns the case when the number of parameters in the $A_p(q)$, $B_p(q)$ and $C_p(q)$ polynomials are equal. This is a typical restriction of the SID methods when used for estimation of linear parametric models:

Assume that one would like to find an ARMAX model, given a set of data (2), with four parameters in $B_p(q)$ but only two parameters in $A_p(q)$. Using a SID method, this would lead to the estimation of a fourth order state-space model, which in its turn will yield four parameters in the $A_p(q)$ polynomial. Ideally, two of these parameters would then be equal to zero, which means that two of the poles will be located in the origin. In the existing SID methods, there is no widely known way to restrict the location of the poles when estimating the system matrix $A(\theta)$ in (3), and thus the extra parameters in the $A_p(q)$ polynomial will generally be nonzero. The main goal of this paper is to find tools that can work around these kinds of problems when using the SID methods.

Even though the main focus of this paper is on the discrete-time problem formulation, the continuous-time case follows readily if one substitutes the use of a discrete-time SID method for a continuous-time equivalent, such as the ones presented in e.g. Van Overschee and De Moor (1996a); McKelvey (1995).

In Section 2, the model structures used in this paper are presented and in Section 3, a short review of the subspace method is given. In Sections 4 and 5, the proposed algorithm is described for the OE and ARMAX model structures, respectively. In Section 6 the ideas are extended to certain linear gray-box models and the proposed methods are then validated in Section 7. Finally, some conclusions and future work are outlined in Section 8.

2. MODEL STRUCTURES

The main focus of this paper is the parameter estimation for certain discrete-time linear black-box models.

The ARMAX model structure is defined by

$$A_p(q)y(t) = B_p(q)u(t) + C_p(q)e(t), \tag{4a}$$

with

$$A_p(q) = 1 + a_1q^{-1} + \cdots + a_{n_a}q^{-n_a}, \tag{4b}$$
$$B_p(q) = b_1q^{-n_k} + \cdots + b_{n_b}q^{-(n_k+n_b-1)}, \tag{4c}$$
$$C_p(q) = 1 + c_1q^{-1} + \cdots + c_{n_c}q^{-n_c}, \tag{4d}$$

where q is the shift operator, i.e., $qu(t) = u(t+1)$.

The OE model structure is a special case of (4) with $C_p(q) = A_p(q) \triangleq F_p(q)$, i.e.,

$$y(t) = \frac{B_p(q)}{F_p(q)} u(t) + e(t). \tag{5}$$

In system identification, one is usually interested in estimating the parameters θ (the vector of the unknown coefficients) in (4) or (5) from given input and output data (2). Several methods exist and further information can be found in e.g. Ljung (1999); Söderström and Stoica (1989); Verhaegen and Verdult (2007).

The model structures (4) and (5) can be written in the state-space innovation form

$$x(t+1) = A(\theta)x(t) + B(\theta)u(t) + K(\theta)e(t), \tag{6a}$$
$$y(t) = C(\theta)x(t) + D(\theta)u(t) + e(t), \tag{6b}$$

where the system matrices depend linearly on the parameters. A special case of (6), which will play an important part in methods that follow, is the *Observer Canonical Form* (OCF)

$$x(t+1) = \begin{pmatrix} -a_1 & 1 & 0 & \dots & 0 \\ -a_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n-1} & 0 & 0 & \dots & 1 \\ -a_n & 0 & 0 & \dots & 0 \end{pmatrix} x(t) + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{pmatrix} u(t) + \begin{pmatrix} c_1 - a_1 \\ c_2 - a_2 \\ \vdots \\ c_{n-1} - a_{n-1} \\ c_n - a_n \end{pmatrix} e(t), \tag{7a}$$

$$y(t) = (1\ 0\ 0\ \dots\ 0)\, x(t) + e(t), \tag{7b}$$

which has a direct correspondence to the ARMAX model given in (4). The unknown parameters in the system matrices (7) correspond directly to the coefficients of the polynomials (4) when $n_k = 1$. For the OE model structure (5), there is no process noise, and thus no matrix $K(\theta)$ in (6) to estimate.

3. SUBSPACE IDENTIFICATION

Assume now that a set of input and output data (2) is given. A solution to the state-space equation (6) can be written as

$$Y = \Gamma X + HU + V, \tag{8}$$

see e.g. Van Overschee and De Moor (1996b). Here U, Y, and V are the Hankel matrices for the input, the output and the noise, respectively, Γ is the extended observability matrix, H is the Markov parameter matrix, and X is the collection of the state vectors for the different measurements.

The basic idea of the subspace method is that, by finding an appropriate projection matrix $\Pi_{U^T}^{\perp}$ and instruments Φ, one can get rid of the last two terms on the right hand side of (8). It can then be shown that the column space of the extended observability matrix Γ is given by the column space of $Y\Pi_{U^T}^{\perp}\Phi^T$. Finding Γ is the most critical step in the subspace algorithm, and several design choices need to be made. Also, it is in this step the discrete-time and continuous-time algorithms differ the most.

Once Γ has been found, it is straight-forward to find estimates $\widehat{A}$ and $\widehat{C}$ of the matrices $A(\theta)$ and $C(\theta)$ in (6). Asymptotically, as the number of data points goes to infinity, it holds that

$$\widehat{A} = TA(\theta)T^{-1}, \quad \widehat{C} = C(\theta)T^{-1}, \tag{9}$$

for some unknown similarity transform T.

The next step is to calculate the estimates of $B(\theta)$ and $D(\theta)$, which is done by solving the least-squares problem

$$\min_{\theta} \frac{1}{N} \sum_{t=0}^{N-1} \|y(t) - \varphi(t)^T \eta\|_2^2. \tag{10}$$

with

$$\varphi(t)^T = \left(\left(\sum_{\tau=0}^{t-1} u(\tau)^T \otimes \widehat{C}\widehat{A}^{t-\tau-1} \right) \quad \left(u(t)^T \otimes I_{n_y} \right) \right)$$

where $\otimes$ is the Kronecker product, vec-operator stacks the columns of the argument in a row vector, and

$$\eta = \begin{pmatrix} \operatorname{vec} B \\ \operatorname{vec} D \end{pmatrix}.$$

The final step, which is optional, is to estimate $K(\theta)$ in (6). To this end, the state sequence X in (8) needs to be reconstructed, which can be done as described in Van Overschee and De Moor (1996b). Once the innovation covariances have been estimated from the reconstructed state sequence and the estimates of the system matrices, one can find an estimate of $K(\theta)$ as the solution to a Riccati equation. For further details of the algorithms and the design variables, see e.g. Van Overschee and De Moor (1996a,b); Verhaegen and Verdult (2007); Ljung (1999).

4. OE MODELS

In this section and the following, the SID algorithm, described above, will be modified to handle the identification of general OE and ARMAX models. The method for OE models is described first, since it constitutes an intermediate step when estimating an ARMAX model. To avoid notational difficulties, only *Single-Input-Single-Output* (SISO) systems are initially discussed.

4.1 The Simple Case

Assume now that an OE model (5) with $n_b < n_a$ is to be estimated from a given set of data (2). The basic idea of the proposed method is outlined in Algorithm 1.

To illustrate the algorithm, a simple example is given.

Example 1. Assume that an OE model with $n_a = 3$ and $n_b = 2$ is to be estimated, i.e.

$$y(t) = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3}} u(t) + e(t).$$

This equation can be written in the OCF (7) with

$$x(t+1) = \begin{pmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{pmatrix} x(t) + \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} u(t).$$

Here, we are going to use the language of MATLAB and *System Identification Toolbox* (SITB). Algorithm 1 yields the following identification scheme

(1) Find an estimate $\widehat{A}$ of order $n_a = 3$ by some standard SID method. This can be done in MATLAB as:

Algorithm 1 OE.

Input: A set of data (2). The orders n_a and n_b of the $A_p(q)$ and $B_p(q)$ polynomials with $n_b < n_a$.

(1) Find estimates $\widehat{A}$ and $\widehat{C}$ of the $A(\theta)$ and $C(\theta)$ matrices of order n_a with some standard SID method.
(2) Determine the OCF $\widehat{A}_O$ and $\widehat{C}_O$ of $\widehat{A}$ and $\widehat{C}$ by determining the coefficients of the characteristic polynomial of $\widehat{A}$.
(3) Estimate the $B(\theta)$ matrix via the least-squares problem (10) with equality constraints on any known elements using the estimates $\widehat{A}_O$ and $\widehat{C}_O$, see e.g. Golub and Van Loan (1996).
(4) Identify the polynomials $F_p(q)$ and $B_p(q)$ in (5) with the coefficients of the characteristic polynomial of $\widehat{A}_O$ and the elements of $\widehat{B}_O$, respectively.

```
m = n4sid(data,na); A = get(m,'a');
```

(2) Transform the matrices to OCF by deriving the characteristic polynomial of the system matrix. This can be done in MATLAB as:

```
a = poly(A);
Ao = [-a(2:end).',eye(na,na-1)];
Co = eye(1,na);
```

(3) Estimate the matrix B from (10) with equality constraints. If (10) is rewritten as a quadratic optimization problem

$$\min_\theta \frac{1}{2}\theta^T H\theta + f^T\theta,$$

then one can solve the equality constrained problem in MATLAB (using the *Optimization Toolbox*) as:

```
Bo = quadprog(H,f,[],[],Aeq,beq),
```

where the constraint $A_{\text{eq}}\theta = b_{\text{eq}}$ is defined by

```
Aeq = [0,0,1]; beq = 0;
```

(4) The polynomials $F_p(q)$ and $B_p(q)$ in (5) are now found in the first column of the estimated A and B matrices, respectively.

Note that the MATLAB code used in the example is not optimized in any way, it is merely used to illustrate the different steps in the proposed algorithm in a simple manner. For example, in the first step of the algorithm, only the $A(\theta)$ matrix needs to be estimated, while the `n4sid` command estimates a full state-space model.

4.2 The General Case

Let us turn to the case when $n_b \geq n_a$ as discussed in Section 1. Instead of presenting general formulas, which are quite lengthy and non-instructive, a simple but fairly general example will be used to illustrate the ideas.

Example 2. Let $n_b = 5$, $n_a = 2$ and $n_k = 3$, then (5) can be written as

$$y(t) = \frac{b_1+b_2q^{-1}+b_3q^{-2}+b_4q^{-3}+b_5q^{-4}}{1+a_1q^{-1}+a_2q^{-2}}q^{-3}u(t)+e(t).$$

Rewriting the above equation to OCF (7) with delayed input yields

$$x(t+1) = \begin{pmatrix} -a_1 & 1 & 0 & 0 & 0 \\ -a_2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} x(t) + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} u(t-2).$$

This means that if one wants to estimate this model, given data, using a SID method, then one needs to estimate a fifth order state-space model. Also, one needs to be able to constrain some of the coefficients in the characteristic polynomial to be zero. To the authors knowledge, no existing subspace method can handle such constraints.

A possible way to solve this problem would be to estimate the $A(\theta)$ matrix, transform it to OCF and just truncate the characteristic polynomial by setting unwanted coefficients to zero. This might work well in some cases, but is probably a bad idea in the case of under-modeling.

Here, we are going to propose a different solution, by introducing *virtual inputs*. Let us rewrite the original equation by splitting the rational expression into two separate terms

$$\begin{aligned} y(t) =& \frac{b_1+b_2q^{-1}+b_3q^{-2}+b_4q^{-3}+b_5q^{-4}}{1+a_1q^{-1}+a_2q^{-2}}q^{-3}u(t)+e(t) \\ =& \frac{b_1+b_2q^{-1}+b_3q^{-2}}{1+a_1q^{-1}+a_2q^{-2}}q^{-3}u(t) \\ &+ \frac{b_4+b_5q^{-1}}{1+a_1q^{-1}+a_2q^{-2}}q^{-6}u(t) + e(t). \end{aligned}$$

Now, polynomial division of the rational expressions yields

$$\begin{aligned} y(t) =& \left(b_1 + \frac{(b_2-b_1a_1)q^{-1}+(b_3-b_1a_2)q^{-2}}{1+a_1q^{-1}+a_2q^{-2}}\right)u(t-3) \\ &+ \left(b_4 + \frac{(b_5-b_4a_1)q^{-1}+(-b_4a_2)q^{-2}}{1+a_1q^{-1}+a_2q^{-2}}\right)u(t-6), \end{aligned}$$

which can be written as

$$\begin{aligned} x(t+1) =& \begin{pmatrix} -a_1 & 1 \\ -a_2 & 0 \end{pmatrix} x(t) + \begin{pmatrix} b_2-b_1a_1 & b_5-b_4a_1 \\ b_3-b_1a_2 & -b_4a_2 \end{pmatrix} \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}, \\ y(t) =& (1\ 0)\, x(t) + (b_1\ b_4) \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} + e(t), \end{aligned}$$

where $u_1(t) \triangleq u(t-3)$ and $u_2(t) \triangleq u(t-6)$. Thus, the original model can now be estimated as a second order state-space model with two input signals instead of one. It follows that the constraints on the characteristic polynomial have been implicitly taken care of.

Since the estimate of $A(\theta)$ is found in the first step of the algorithm and transformed to the OCF in the second step, linear equality constraints can now be used when estimating $B(\theta)$ and $D(\theta)$ via (10), as shown in the previous section. In this particular example, the constraint on the least-squares problem (10) is given by

$$(0\ 0\ 0\ 1\ 0\ -\hat{a}_2) \begin{pmatrix} \operatorname{vec} B \\ \operatorname{vec} D \end{pmatrix} = 0.$$

Even though the example is quite simple, all the difficulties that can be encountered in the general case are present. □

From the simple example above, one can see how the general algorithm works and the only difference from Algorithm 1 is that the third step now includes the estimation of the $D(\theta)$ matrix with equality constraints.

Now, let us turn to the ARMAX case.

5. ARMAX MODELS

To estimate the full ARMAX model (4), the only thing that remains is to estimate the $K(\theta)$ matrix in (6), which has the form (7). Here, we must have $n_c \leq n_a$, which limits the flexibility of the noise model.

To be able to find an estimate of the matrix $K(\theta)$, one needs to estimate the process and measurement noises. To this end, one needs to reconstruct the state matrix X in (8), which can be done as described in Van Overschee and De Moor (1996b) page 169. Once the estimates of the system matrices and the state sequence in (8) have been found, the noise can be estimated as

$$\begin{pmatrix} \widehat{W} \\ \widehat{V} \end{pmatrix} = \begin{pmatrix} \overline{X} \\ \underline{Y} \end{pmatrix} - \begin{pmatrix} \widehat{A} & \widehat{B} \\ \widehat{C} & \widehat{D} \end{pmatrix} \begin{pmatrix} \underline{X} \\ \underline{U} \end{pmatrix}, \tag{11}$$

where the overline means removing the first column and the underline removing the last column. From (6) it holds that

$$K(\theta)\widehat{V} = \widehat{W}. \tag{12}$$

A least-squares estimate of $K(\theta)$ via (12) can now be found, where we add equality constraints on any known element. This yields in turn an estimate of the $C_p(q)$ polynomial in (4) as shown in (7). An algorithm for ARMAX model identification is given in Algorithm 2.

Algorithm 2 ARMAX.

Input: A collection of input and output data (2) and integers n_a, n_b, and n_c. Limitation: $n_c \leq n_a$.

(1) Find estimates $\widehat{A}$, $\widehat{B}$, $\widehat{C}$ and $\widehat{D}$ by Algorithm 1.
(2) Reconstruct the state matrix X, which can be done as in Van Overschee and De Moor (1996b) page 169.
(3) Find $\widehat{W}$ and $\widehat{V}$ from (11).
(4) Estimate $K(\theta)$ by solving (12) in a least-squares sense with equality constraints on any known elements.
(5) Transform the estimated state-space model to the polynomial form (4).

Let us consider a small example:

Example 3. Assume that one wants to estimate an ARMAX model

$$y(t) = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}} u(t) + \frac{1 + c_1 q^{-1}}{1 + a_1 q^{-1} + a_2 q^{-2}} e(t).$$

The equation above can be written as

$$x(t+1) = \begin{pmatrix} -a_1 & 1 \\ -a_2 & 0 \end{pmatrix} x(t) + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} u(t) + \begin{pmatrix} c_1 - a_1 \\ -a_2 \end{pmatrix} e(t).$$

Thus, there is only one linear constraint when estimating the matrix $K(\theta)$ in the third step of Algorithm 2, i.e. the last element of $K(\theta)$ should be equal to $-a_2$. □

The estimation procedure of the ARMAX model structure is thus a simple extension of the OE algorithm given in Section 4.

6. SPECIAL GRAY-BOX MODELS

The ideas presented so far have been based on a coordinate change of the state-space model to the OCF, which works well for certain black-box model structures. But what if a different structure is present in the state-space form, like the ones in linear gray-box models? Is there anything that can be done in such cases?

Consider the relation between the estimates $\widehat{A}$ and $\widehat{C}$ and their corresponding gray-box representation $A(\theta)$ and $C(\theta)$, respectively, which has a certain structure that we would like to preserve, i.e.

$$\widehat{A} = TA(\theta)T^{-1}, \quad \widehat{C} = C(\theta)T^{-1},$$

for some similarity transform T. The expression above can be rewritten as

$$\widehat{A}T = TA(\theta), \quad \widehat{C}T = C(\theta),$$

and vectorized into

$$\begin{pmatrix} I \otimes \widehat{A} - A(\theta)^T \otimes I \\ I \otimes \widehat{C} \end{pmatrix} \operatorname{vec} T = \begin{pmatrix} 0 \\ \operatorname{vec} C(\theta) \end{pmatrix}. \tag{13}$$

For certain linear gray-box structures, where all the unknown parameters of $A(\theta)$ lie in one row or one column and all the elements of $C(\theta)$ are known, the equation (13) might have a solution with respect to T. These kind of structures were also noted to be favorable in Xie and Ljung (2002).

Example 4. Let us consider (13) in the OCF case (7). Then the left hand side of (13) has the form

$$\begin{pmatrix} \widehat{A} + a_1 I & a_2 I & a_3 I & \dots & a_{n_a-1} I & a_{n_a} I \\ -I & \widehat{A} & 0 & \dots & 0 & 0 \\ 0 & -I & \widehat{A} & \dots & 0 & 0 \\ 0 & 0 & -I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -I & \widehat{A} \\ \widehat{C} & 0 & 0 & \dots & 0 & 0 \\ 0 & \widehat{C} & 0 & \dots & 0 & 0 \\ 0 & 0 & \widehat{C} & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \widehat{C} & 0 \\ 0 & 0 & 0 & \dots & 0 & \widehat{C} \end{pmatrix},$$

and the right hand side is a vector with all elements equal to zero except the $n_a^2 + 1$ element that is equal to one.

Thus, it is only the first n_a equations that depend on the unknown parameters. The remaining number of equations sum up to n_a^2, which is equal to the number of unknowns in T. Thus, (13) is solvable for T whenever the lower part of the matrix above has full rank. □

So, how does one proceed? In Van Overschee and De Moor (1996b), the N4SID subspace algorithm is formulated as a least-squares problem. It begins by reconstructing the state matrix X from the Hankel matrices Y and U. Now, the system matrices can be estimated by solving the least-squares problem

$$\underset{A,B,C,D}{\arg\min} \left\| \begin{pmatrix} \overline{X} \\ \underline{Y} \end{pmatrix} - \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \underline{X} \\ \underline{U} \end{pmatrix} \right\|_F^2, \tag{14}$$

where the overline means removing the first column and the underline removing the last column.

This formulation enables a two step method for estimating the parameters in gray-box models. First, the algorithm is used to estimate the system matrices. The resulting matrices are then used to determine the similarity transform T that takes the system as close as possible to the desired form. Then one transforms the state matrix X with the determined transform. Finally, the system matrices are re-estimated via (14) where one adds equality constraints for the known elements. This method is evaluated in a simulation example in the next section.

7. SIMULATIONS

In this section, some simulation results are given. The method presented in this paper is compared to the results using the *System Identification Toolbox* (SITB) in MATLAB. The SITB is using the IV method as an initial estimator, and then PEM to refine the estimate. Since the modified subspace method is intended to be used as an initial parameter estimator, the PEM iterations in SITB have been turned off to get a comparable result.

7.1 OE Model

In the first example, a Monte Carlo study of the estimation of an OE model, using the method presented in Section 4, is presented.

Example 5. Let the true system be given by (5) with

$$B_p(q) = q^{-2} - 0.4834q^{-3} - 0.2839q^{-4} - 0.02976q^{-5}$$
$$F_p(q) = 1 + 0.7005q^{-1} - 0.2072q^{-2}$$

and let $e(t)$ be white Gaussian noise of unit variance and zero mean, i.e. an OE$(4, 2, 2)$ structure. For the Monte Carlo analysis, output data $y(t)$ have been generated with $M = 1,000$ realizations of a zero mean white Gaussian process, with unit variance and length $1,000$, as input $u(t)$. The noise also has a different realization for each Monte Carlo run.

The result is given in Table 1. The results using the proposed subspace method are denoted by SID and IV denotes the results using SITB.

Table 1. Monte Carlo analysis of the BIAS (16) and the VAR (17) when estimating the parameters in the model (15).

	BIAS		VAR	
	SID	IV	SID	IV
b_1	0.0026	0.0233	0.0013	0.2223
b_2	0.0667	0.1660	0.2266	176.9809
b_3	0.1055	0.0684	0.0648	15.3032
b_4	0.0862	0.0810	0.0739	29.9095
a_1	0.0642	0.0491	0.2275	0.2035
a_2	0.0042	0.0455	0.1808	0.1777

The first two columns, BIAS, gives a measure of the bias of the parameter estimates:

$$\widehat{\text{Bias}}\,\hat{\theta} \triangleq \frac{1}{M}\Big|\sum_{i=1}^{M}(\theta_0 - \hat{\theta}_i)\Big|. \tag{16}$$

The last two columns, VAR, present the variance of the parameter estimate

$$\widehat{\text{Var}}\,\hat{\theta} \triangleq \frac{1}{M-d}\sum_{i=1}^{M}(\hat{\theta}_i - \bar{\hat{\theta}}_i)^2, \tag{17}$$

where d is the number of parameters and the bar denotes the mean value.

For this example, the SID method seems to yield lower variance when estimating the $B_p(q)$ polynomial than the traditional IV estimator. On the other hand, the IV estimator yields lower bias than the proposed SID method for some of the parameters. □

7.2 ARMAX Model

In the next example, the method presented in Section 5 is evaluated as a method for estimating ARMAX models

Example 6. Let the true system be given by (4) with

$$A_p(q) = 1 - 0.06353q^{-1} + 0.006253q^{-2} + 0.0002485q^{-3}$$
$$B_p(q) = 1 - 0.8744q^{-1} - 0.3486q^{-2} + 0.331q^{-3}$$
$$C_p(q) = 1 + 0.3642q^{-1}$$

and let $e(t)$ be white Gaussian noise of unit variance and zero mean, i.e. an ARMAX$(3, 4, 1, 0)$ structure. For the Monte Carlo analysis, output data $y(t)$ have been generated with $M = 1,000$ realizations of a zero mean white Gaussian process, with unit variance and length $1,000$, as input $u(t)$. The noise also has a different realization for each Monte Carlo run.

The result is given in Table 2. For this choice of system, the IV and the SID method yield similar results. The similarity is probably due to the length of the data set, which helps the IV in terms of variance. For shorter data sets, like the one in Example 5, the SID is usually significantly better in terms of variance but gains some bias. □

Table 2. Monte Carlo analysis of the BIAS (16) and the VAR (17) when estimating the parameters in the model (18).

	BIAS		VAR	
	SID	IV	SID	IV
a_1	0.0137	0.0329	0.0101	0.0447
a_2	0.0025	0.0100	0.0016	0.0047
a_3	0.0040	0.0072	0.0004	0.0021
b_1	0.0000	0.0001	0.0001	0.0001
b_2	0.0137	0.0367	0.0102	0.0560
b_3	0.0083	0.0190	0.0027	0.0153
b_4	0.0018	0.0157	0.0036	0.0125
c_1	0.0141	0.0141	0.0102	0.0213

7.3 Gray-box Model

In the next example, the two-step method described in Section 6 is used to estimate a gray-box model.

Example 7. Let the true system be given by

$$x(t+1) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \theta_1 & 0 & \theta_2 \end{pmatrix} x(t) + \begin{pmatrix} \theta_3 \\ 0 \\ \theta_4 \end{pmatrix} u(t) \tag{19a}$$

$$y(t) = (1\ 0\ 0)\, x(t) + e(t) \tag{19b}$$

with $\theta = (-0.3, 0.3, 0.1, -0.1)^T$. Here, $e(t)$ is white Gaussian noise of unit variance. For the Monte Carlo analysis, output data $y(t)$ have been generated with $M = 500$ realizations of a zero mean white Gaussian process, with unit variance and length $10,000$, as input $u(t)$. The parameters θ have been estimated for each realization with the two-step method presented in Section 6. The mean value and the variance of the parameter estimates have been given by

$$\bar{\hat{\theta}} = (-0.2873\ \ 0.2675\ \ 0.1003\ \ -0.0921)^T, \tag{20a}$$

$$\widehat{\text{Var}}\,\hat{\theta} = 10^{-2} \times (0.2395\ \ 0.6518\ \ 0.0096\ \ 0.0144)^T. \tag{20b}$$

If no consideration is given to the zero in between the two parameters θ_1 and θ_2 in the system matrix $A(\theta)$ in (19), the mean values of parameter estimates become

$$\bar{\hat{\theta}}_1 = -0.2049, \quad \bar{\hat{\theta}}_2 = 0.3957, \tag{21a}$$

and their variances

$$\widehat{\mathrm{Var}}\,\tilde{\theta}_1 = 0.0293, \quad \widehat{\mathrm{Var}}\,\tilde{\theta}_2 = 0.1004. \tag{21b}$$

This shows the increase in variance when not taking the zero between the two unknown parameters in $A(\theta)$ into account. Thus, the two-step method is a candidate for initial parameter estimation of certain gray-box models, when the structure of the matrices is appropriate. □

7.4 Continuous-time Model

Finally, let us evaluate the proposed method when estimating a continuous-time OE model from discrete-time data. The subspace method used is presented in McKelvey (1995); Van Overschee and De Moor (1996a).

Example 8. Let the true system be given by the Rao and Garnier (2002) test system

$$y(t) = \frac{-6400p+1600}{p^4+5p^3+408p^2+416p+1600}u(t),$$

where p is the ordinary differential operator. This system is non-minimum phase, it has one fast oscillatory mode with relative damping 0.1 and one slow mode with relative damping 0.25, which makes it a difficult system to identify.

For data collection, the system is sampled with a rate of $T_s = 10\,\mathrm{ms}$. The sampled system is then simulated with a random binary input of length 4,095 and white Gaussian noise of unit variance is then added to the output, which yields a signal to noise ratio of 4.9 dB.

The validation of the proposed method is given as a Bode plot in Figure 1 and cross validation on validation data is given in Figure 2. The results are promising even in the continuous-time case.

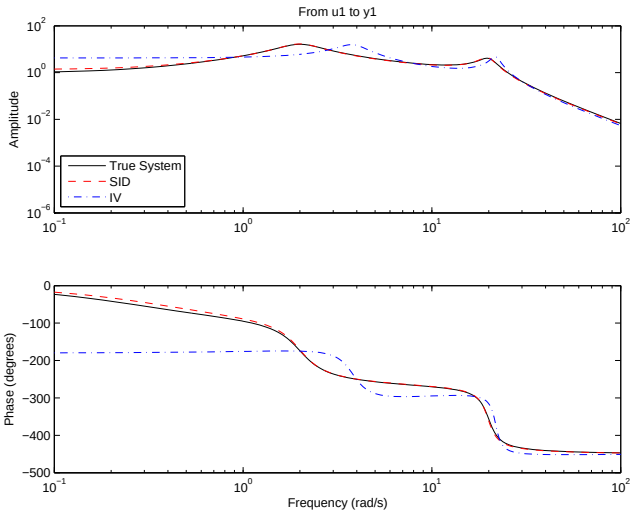

Fig. 1. The Bode plot of the true system and the estimates given by the SID and the IV methods.

8. CONCLUSIONS AND FUTURE WORK

A new algorithm for initial parameter estimation of certain linear model structures has been presented, which makes use of the standard SID methods. Even though only the SISO case is presented, the method is also valid for MIMO systems. This follows from the use of the OCF which can handle MISO systems, and MIMO system can be considered

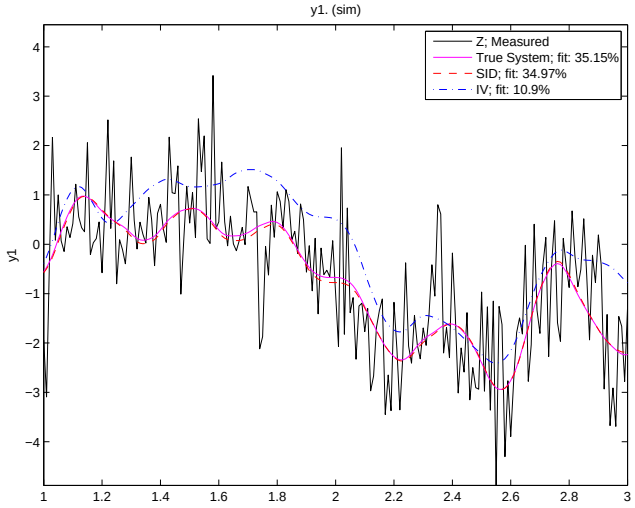

Fig. 2. The fit of the true and the estimated models on validation data.

as several MISO systems. The method is also valid for both discrete-time and continuous-time identification.

The proposed method might be generalized to handle Box-Jenkins models, which is an topic for future work. Also, the proposed algorithms might enable automatic order selection of the model structures treated in this paper, if merged with some method for order selection of linear regression models.

ACKNOWLEDGEMENTS

The authors would like to thank the Swedish Research Council and the Vinnova Industry Excellence Center LINKSIC for support.

REFERENCES

Bauer, D. (2005). Estimating linear dynamical systems using subspace methods. *Econometric Theory*, 21, 181–211.

Golub, G. and Van Loan, C. (1996). *Matrix Computations*. Johns Hopkins University Press, 3rd edition.

Ljung, L. (1999). *System Identification, Theory for the User*. Prentice Hall, 2nd edition.

McKelvey, T. (1995). *Identification of State-Space Models from Time and Frequency Data*. Ph.D. thesis, Linköpings Universitet, Linköping, Sweden.

Rao, G.P. and Garnier, H. (2002). Numerical illustrations of the relevance of direct continuous-time model identification. In *Proceedings of the 15th IFAC World Congress*. Barcelona, Spain.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice Hall.

Van Overschee, P. and De Moor, B. (1996a). Continuous-time frequency domain subspace system identification. *Signal Processing*, 52(5), 179–194.

Van Overschee, P. and De Moor, B. (1996b). *Subspace Identification for Linear Systems*. Kluwer Academic Publishers.

Verhaegen, M. and Verdult, V. (2007). *Filtering and System Identification*. Cambridge University Press.

Xie, L. and Ljung, L. (2002). Estimate physical parameters by black-box modeling. In *Proceedings of the 21st Chinese Control Conference*, 673–677.

Avdelning, Institution
Division, Department

Division of Automatic Control
Department of Electrical Engineering

Datum
Date

2009-04-03

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English

☐ _____

Rapporttyp
Report category

☐ Licentiatavhandling
☐ Examensarbete
☐ C-uppsats
☐ D-uppsats
☒ Övrig rapport
☐ _____

ISBN
—

ISRN
—

Serietitel och serienummer
Title of series, numbering

ISSN
1400-3902

LiTH-ISY-R-2890

URL för elektronisk version

`http://www.control.isy.liu.se`

Titel Handling Certain Structure Information
Title in Subspace Identification

Författare Christian Lyzell, Martin Enqvist, Lennart Ljung
Author

Sammanfattning
Abstract

The prediction-error approach to parameter estimation of linear models often involves solving a non-convex optimization problem. In some cases, it is therefore difficult to guarantee that the global optimum will be found. A common way to handle this problem is to find an initial estimate, hopefully lying in the region of attraction of the global optimum, using some other method. The prediction-error estimate can then be obtained by a local search starting at the initial estimate. In this paper, a new approach for finding an initial estimate of certain linear models utilizing structure and the subspace method is presented. The polynomial models are first written on the observer canonical state-space form, where the specific structure is later utilized, rendering least-squares estimation problems with linear equality constraints.

Nyckelord
Keywords System identification, Subspace method, Black-box models.