ON- AND OFF-LINE IDENTIFICATION OF LINEAR STATE SPACE MODELS

Marc Moonen*,Bart De Moor, Lieven Vandenberghe*, Joos Vandewalle
ESAT Katholieke Universiteit Leuven
K.Mercierlaan 94, 3030 Heverlee, Belgium
tel 016/22 09 31 telex 25941 elekul fax 32/16/221855

Abstract

A geometrically inspired matrix algorithm is derived for the identification of state space models for multivariable linear time-invariant systems using (possibly noisy) input-output measurements only. As opposed to other –mostly stochastic– identification schemes, no variance-covariance information whatever is involved, and only a limited number of I/O-data are required for the determination of the system matrices.

Hence, the algorithm can be best described and understood in the matrix formalism, and consists in the following two steps: First a state vector sequence is realized as the intersection of the row spaces of two block Hankel matrices, constructed with I/O-data. Then the system matrices are obtained at once from the least squares solution of a set of linear equations.

When dealing with noisy data, this algorithm draws its excellent performance from repeated use of the numerically stable and accurate singular value decomposition Also, the algorithm is easily applied to slowly time-varying systems using windowing or exponential weighting. These results are illustrated by examples, including the identification of an industrial plant.
Keywords : multivariable systems, system identification, singular value decomposition.

1 Introduction

Identification aims at finding a mathematical model from the measurement record of inputs and outputs of a system. A state space model is a most obvious choice for a mathematical representation because of its widespread use in system theory and control. Still, reliable general purpose state space identification schemes have not become standard tools so far, mostly due to the computational complexity involved (Ho and Kalman 1965, Kung 1978, Zeiger and Mc Ewen 1974).

*Supported by the N.F.W.O. (Belgian National Fund for Scientific Research)

The theory of canonical correlation analysis, independently developed in the midthirties by Hotelling (Hotelling 1936) and Obukhov (the idea of using SVD to compute the principal angles and vectors being due to Bjorck and Golub (Golub and Van Loan 1983), has been intensively applied to the stochastic identification problem, where as a major departure canonical variate analysis is used to choose linear combinations of the past of the random process to optimally predict the future of the process. The analysis of a system in terms of past and future naturally leads to a state space description (Akaike 1974, Akaike 1975, Baram 1981, Ramos and Verriest 1984, Larimore 1984). Nevertheless, the intensive use of covariance information is a major drawback when it comes to practice, since finite data records reveal only poor approximations for covariance matrices.

In this paper, a novel approach is presented, that shows much resemblance to the canonical variate methods, but no variance-covariance information whatsoever is involved, and only a finite number of I/O-data are required for the determination of the system matrices. The main step in the identification procedure consists in the singular value decomposition of a block Hankel matrix, constructed with I/O-data. As it will turn out that only the left singular basis is required, both the computational load and the noise sensitivity are considerably reduced. Moreover, the identification scheme is easily converted into an adaptive version. In section 2, useful properties of dynamic systems are briefly described, which are used in section 3 to show how a sequence of state vectors can be calculated. The system matrices are then identified by solving an overdetermined set of linear equations (Section 4). The off-line algorithm is summarized in section 5, and converted into an adaptive on-line algorithm for slowly time-varying systems in section 6. Both strategies are illustrated by examples.

2 Dynamic systems

The most general linear discrete-time multivariable state space model can be written as

$$\begin{aligned} x[k+1] &= A_k.x[k] + B_k.u[k] + w[k] \\ y[k] &= C_k.x[k] + D_k.u[k] + v[k] \end{aligned} \tag{1}$$

where $u[k], y[k]$ and $x[k]$ denote the input (m-vector), output (l-vector) and state vector at time k, the dimension of $x[k]$ being the minimal system order n. A_k, B_k, C_k and D_k are the unknown system matrices at time k to be identified, making use only of recorded I/O-sequences $u[k], u[k+1], \ldots$ and $y[k], y[k+1], \ldots$ As it is obvious that only the observable part of the system can be identified from observed I/O-data, it can be assumed that the system is completely observable, thus omitting the unobservable part at the very outset.

$w[k]$ and $v[k]$ are additional unknown noise-sequences, accounting for measurement noise, process noise, model mismatch, etc. They will be identified as the residuals of the set of equations that determine the system matrices (section 4), and can thus be omitted for a while. Also, for the time being, we consider only time-invariant systems, so that the state space equations eventually reduce to

$$\begin{aligned} x[k+1] &= A.x[k]+B.u[k] \\ y[k] &= C.x[k]+D.u[k] \end{aligned} \tag{2}$$

We now state two important theorems that will be used throughout the sequel.

Theorem 1 *Sequences u,y,x that satisfy equations (2), also satisfy the following general structured I/O-equation :*

$$Y_h = \Gamma_i.X + H_t.U_h \tag{3}$$

Y_h is a block Hankel matrix (i block rows, j columns) containing the consecutive outputs :

($y[k]$ is a $l \times 1$ vector, where l is the number of outputs)

$$Y_h \quad = \quad \begin{bmatrix} y[k] & y[k+1] & \dots & \dots & \dots & y[k+j-1] \\ y[k+1] & y[k+2] & \dots & \dots & \dots & y[k+j] \\ y[k+2] & y[k+3] & \dots & \dots & \dots & y[k+j+1] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ y[k+i-1] & y[k+i] & \dots & \dots & \dots & y[k+j+i-2] \end{bmatrix}$$

U_h is a block Hankel matrix with the same block dimensions as Y_h, containing the consecutive inputs. ($u[k]$ is a $m \times 1$ vector, where m is the number of inputs)

$$U_h = \begin{bmatrix} u[k] & u[k+1] & \dots & \dots & \dots & u[k+j-1] \\ u[k+1] & u[k+2] & \dots & \dots & \dots & u[k+j] \\ u[k+2] & u[k+3] & \dots & \dots & \dots & u[k+j+1] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u[k+i-1] & u[k+i] & \dots & \dots & \dots & u[k+j+i-2] \end{bmatrix}$$

X contains consecutive state vectors :

$$X = [x[k]\ x[k+1]\ x[k+2]\ \dots\ x[k+j-1]]$$

Γ_i is an extended observability matrix :

$$\Gamma_i = \begin{bmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{i-1} \end{bmatrix}$$

Finally H_t is a lower triangular block Toeplitz matrix containing the Markov parameters :

$$H_t = \begin{bmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ CA^2B & CAB & CB & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{bmatrix}$$

Proof : straightforward by repeated substitution of equations (2).

□

Instead of going into details, we loosely state that i and j should be chosen "sufficiently large" (so that Y_h and U_h contain enough information on the system), and in particular $j \gg \max(mi, li)$ ("very rectangular" block Hankel matrices), as this will reduce both the computational load and the noise sensitivity (see below).

Theorem 2 *Let Y_h, U_h and X be defined as in the previous theorem, and let H denote the concatenation of Y_h and U_h :*

$$H = \begin{bmatrix} Y_h \\ U_h \end{bmatrix}$$

then, under the conditions that
1. $\mathrm{rank}(X) = n$ *, i.e. all modes are sufficiently excited (n being the minimal system order), and*
2. $\mathrm{span}_{\mathrm{row}}(X) \cap \mathrm{span}_{\mathrm{row}}(U_h) = \emptyset$,
the following rank property holds :

$$\mathrm{rank}(H) = \mathrm{rank}(U_h) + n \tag{4}$$

Also, when
3. $\mathrm{rank}(U_h) = mi =$*number of rows in* U_h ,
this rank property reduces to

$$\mathrm{rank}(H) = mi + n \tag{5}$$

Proof
From equation (3) it follows that :

$$Y_h.U_h^{\perp} = \Gamma_i.X.U_h^{\perp}$$

and then of course :

$$\mathrm{rank}(Y_h.U_h^{\perp}) = \mathrm{rank}(\Gamma_i.X.U_h^{\perp})$$

where the columns of $U_h^{\perp}$ span the kernel of U_h (not trivial since $j \gg mi$).
Since Γ_i has full column rank (cfr. observability) :

$$\begin{aligned} \mathrm{rank}(\Gamma_i.X.U_h^{\perp}) &= \mathrm{rank}(X.U_h^{\perp}) - \dim(\mathrm{span}_{\mathrm{col}}(X.U_h^{\perp}) \cap (\mathrm{span}_{\mathrm{row}}(\Gamma_i))^{\perp}) \\ &= \mathrm{rank}(X.U_h^{\perp}) - \dim(\mathrm{span}_{\mathrm{col}}(X.U_h^{\perp}) \cap \emptyset) \\ &= \mathrm{rank}(X.U_h^{\perp}) \end{aligned}$$

By making use of condition 2 :

$$\begin{aligned} \mathrm{rank}(X.U_h^{\perp}) &= \mathrm{rank}(X) - \dim(\mathrm{span}_{\mathrm{row}}(X) \cap (\mathrm{span}_{\mathrm{col}}(U_h^{\perp})^{\perp}) \\ &= \mathrm{rank}(X) - \dim(\mathrm{span}_{\mathrm{row}}(X) \cap \mathrm{span}_{\mathrm{row}}(U_h)) \\ &= \mathrm{rank}(X) \end{aligned}$$

Finally, under condition 1 :

$$\text{rank}(X) = n$$

By combining all the above equations, one obtains

$$\text{rank}(Y_h.U_h^{\perp}) = n$$

and this, in fact, means that the row space of Y_h adds n dimensions to the row space of U_h, which proves equation (4).

$\square$

This theorem allows us to estimate the system order, prior to further identification of the system matrices.

Note on condition 1 : $\text{rank}(X) = n$, in other words all modes should be sufficiently excited (persistant excitation). When certain modes are not, i.e. unobservable in the I/O-data currently under investigation, they cannot be identified either and application of the above rank property will reveal too low a system order, this problem being inherent in system identification.

Note on condition 2 : $\text{span}_{\text{row}}(X) \cap \text{span}_{\text{row}}(U_h) = \emptyset$.
When this condition is not satisfied, $\text{rank}(X.U_h^{\perp}) < \text{rank}(X)$ (rank cancellation), and again application of the rank property will reveal an underestimation of the system order. However it can be experimentally verified that rank cancellation is not generic, and the probability that rank cancellation occurs, decreases for fixed i (number of rows in U_h) with increasing j (number of columns in U_h and X). (In a stochastic framework, this matter would be passed off easily by saying $E(x[k].u[k]^t) = 0$, $E(x[k].u[k+1]^t) = 0$, $\ldots$, where E is the expectation operator.)

Note on condition 3 : $\text{rank}(U_h) = mi =$number of rows in U_h
Similar to the previous ones, this third condition will generically be satisfied when the input is ”sufficiently exciting” (inherent in the identification problem).

In the sequel, it will allways be assumed that these three conditions are satisfied.

3 Determination of a state vector sequence

We now demonstrate how a sequence of state vectors can be calculated as the intersection of the row spaces of two block Hankel matrices, constructed from input-output vectors. Let H_1 and H_2 be the concatenation of Y_{h1}, U_{h1} and Y_{h2}, U_{h2} respectively

$$H_1 = \begin{bmatrix} Y_{h1} \\ U_{h1} \end{bmatrix}, \qquad H_2 = \begin{bmatrix} Y_{h2} \\ U_{h2} \end{bmatrix} \tag{6}$$

where

$$Y_{h1} = \begin{bmatrix} y[k] & y[k+1] & \ldots & \ldots & \ldots & y[k+j-1] \\ y[k+1] & y[k+2] & \ldots & \ldots & \ldots & y[k+j] \\ y[k+2] & y[k+3] & \ldots & \ldots & \ldots & y[k+j+1] \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ y[k+i-1] & y[k+i] & \ldots & \ldots & \ldots & y[k+j+i-2] \end{bmatrix}$$

$$Y_{h2} = \begin{bmatrix} y[k+i] & y[k+i+1] & \dots & \dots & \dots & y[k+i+j-1] \\ y[k+i+1] & y[k+i+2] & \dots & \dots & \dots & y[k+i+j] \\ y[k+i+2] & y[k+i+3] & \dots & \dots & \dots & y[k+i+j+1] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ y[k+2i-1] & y[k+2i] & \dots & \dots & \dots & y[k+2i+j-2] \end{bmatrix}$$

and U_{h1}, U_{h2} similarly constructed. Both matrix pairs satisfy the I/O-equation :

$$Y_{h1} = \Gamma_i.X_1 + H_t.U_{h1} \tag{7}$$
$$Y_{h2} = \Gamma_i.X_2 + H_t.U_{h2} \tag{8}$$

Theorem 3 *If X_2 is defined as*

$$X_2 = [x[k+i]\ x[k+i+1]\ \dots\ x[k+i+j-1]]$$

then

$$\mathrm{span}_{\mathrm{row}}(X2) = \mathrm{span}_{\mathrm{row}}(H_1) \cap \mathrm{span}_{\mathrm{row}}(H_2)$$

(see (6) for a definition of H_1 and $H2$) so that any basis for this intersection constitutes a valid state vector sequence X_2 with the basis vectors as the consecutive row vectors.

Note that different choices for a basis differ in a transformation matrix T that transforms a model A, B, C, D into an equivalent model $T^{-1}.A.T, T^{-1}.B, C.T, D$ (Kailath 1980).

Proof

It is first proven that the dimension of the intersection equals n. Then, the (n-dimensional) row space of X_2 is shown to lie within both row spaces.

By making use of the rank property (5), one derives

$$\dim(H_1) = \dim(H_2) = mi + n$$

where $\dim(M)$ is a shorthand notation for the dimension of M's row space. This rank property holds equally well for the concatenation of H_1 and H_2 :

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$$

$$\dim(H_1 + H_2) = \dim(H) = 2mi + n$$

Applying Grassmann's dimension theorem :

$$\begin{aligned} \dim(H_1 \cap H_2) &= \dim(H_1) + \dim(H_2) - \dim(H_1 + H_2) \\ &= mi + n + mi + n - 2mi - n \\ &= n \end{aligned}$$

From equation (8), one derives

$$X_2 = \Gamma_i^+.Y_{h2} - \Gamma_i^+.H_t.U_{h2} = [\Gamma_i^+ \ -\Gamma_i^+.H_t].\begin{bmatrix} Y_{h2} \\ U_{h2} \end{bmatrix}$$

where Γ_i^+ is Γ_i's pseudo-inverse ($\Gamma_i^+.\Gamma_i = I_{n\times n}$ since Γ_i has full column rank), which shows that X_2's row space lies within H_2's row space. Equally well, X_1's row space lies within H_1's row space. On the other hand, X_1 and U_{h1} completely determine X_2 through

$$X_2 = A^i.X_1 + [A^{i-1}.B \quad \ldots \quad A.B \quad B].U_{h1}$$

and since X_1's row space lies within H_1's row space, the same holds true for X_2's row space.

$\square$

The above theorem allows us to calculate a state vector sequence, making use of measured I/O-data only. Once this state vector sequence is known, the system matrices are easily identified from a set of linear equations, as will be shown in the next section.

In practice, due to perturbations on the measured data (noise, non-linearity, etc.), it occurs that both row spaces do not intersect. An approximate intersection can be calculated though, using the n first principal vectors (canonical variate analysis), n being determined through equation (5).

As it will turn out to be both computationally less demanding and less sensitive to noise on the I/O-data, an alternative procedure is presented : Let the SVD of $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ be

$$H = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \begin{bmatrix} S_{11} & 0 \\ 0 & 0 \end{bmatrix} V^t$$

where the matrices have the following dimensions :

$$\begin{aligned} \dim(U_{11}) &= (mi+li)\times(2mi+n) \\ \dim(U_{12}) &= (mi+li)\times(2li-n) \\ \dim(U_{21}) &= (mi+li)\times(2mi+n) \\ \dim(U_{22}) &= (mi+li)\times(2li-n) \\ \dim(S_{11}) &= (2mi+n)\times(2mi+n) \end{aligned}$$

From

$$[U_{12}^t \; U_{22}^t].\begin{bmatrix} H_1 \\ H_2 \end{bmatrix} = 0$$

or

$$U_{12}^t.H_1 = -U_{22}^t.H_2$$

it follows that the row space of $U_{12}^t.H_1$ equals the required intersection of H_1's and H_2's row spaces. $U_{12}^t.H_1$ contains $2li-n$ row vectors , only n of which are linearly independent (dimension of the intersection). Thus, it remains to select n suitable combinations of these row vectors. One straightforward way would consist in taking the SVD of $U_{12}^t.H_1$ in order to compute a basis for its row space. The following theorem gives the outline of a shortcut to this method, replacing the SVD of $U_{12}^t.H_1$ (a $(2li-n)\times j$-matrix where most of the time j is very large) by a smaller SVD.

Theorem 4 *Let the SVD of* $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ *be*

$$H = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \begin{bmatrix} S_{11} & 0 \\ 0 & 0 \end{bmatrix} V^t$$

then the state vector sequence $X_2 = [x[k+i]\ x[k+i+1]\ \ldots\ x[k+i+j-1]]$ *can be calculated as :*

$$X_2 = U_q^t . U_{12}^t . H_1$$

where U_q *(an* $n \times (2li-n)$ *matrix accounting for the necessary reduction of* $2li-n$ *mutually dependent row vectors of* $U_{12}^t.H_1$ *to n independent vectors) is defined through the SVD of* $U_{12}^t.U_{11}.S_{11}$

$$U_{12}^t.U_{11}.S_{11} = \begin{bmatrix} U_q & U_q^{\perp} \end{bmatrix} \begin{bmatrix} S_q & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_q^t \\ V_q^{\perp t} \end{bmatrix}$$

Proof
Since any basis for the row space of $U_{12}^t.H_1$ is a realization of X_2 (see above), we first calculate its SVD :

$$\begin{aligned} U_{12}^t.H_1 &= U_{12}^t.[U_{11}\ \ U_{12}].\begin{bmatrix} S_{11} & 0 \\ 0 & 0 \end{bmatrix}.V^t \\ &= [U_{12}^t.U_{11}.S_{11}\ \ 0].V^t \\ &= [U_q.S_q.V_q^t\ \ 0].V^t \\ &= U_q.[S_q\ \ 0].(V.V_q)^t \end{aligned}$$

Now, since $U_q^t.U_q = I_{n \times n}$

$$U_q^t.U_{12}^t.H_1 = [S_q\ 0].(V.V_q)^t$$

which is a valuable basis for the row space of $U_{12}^t.H_1$ and thus a realization of X_2.

□

4 Identification of the system matrices

Once X_2 is known, the system matrices can be identified by solving a set of linear equations in a straightforward way :

$$\begin{bmatrix} x[k+i+1] & \ldots & x[k+i+j-1] \\ y[k+i] & \ldots & y[k+i+j-2] \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} x[k+i] & \ldots & x[k+i+j-2] \\ u[k+i] & \ldots & u[k+i+j-2] \end{bmatrix}$$

As this (overdetermined) set of equations should be solved in a least squares sense, the residuals correspond to the noise terms $w[k]$ and $v[k]$ introduced in section 2.

Once again, a computationally more efficient way of computing the system matrices is conceivable, making use of the already calculated SVD of H (concatenation of H_1 and

H_2).The above set of equations can be replaced by a reduced equivalent set, revealing exactly the same least squares solution.

For compact notations, it is useful to first redefine matrices H_1 and H_2 (equation (6)) in the following way :

$$H_1 = \begin{bmatrix} u[k] & u[k+1] & \dots & \dots & \dots & u[k+j-1] \\ y[k] & y[k+1] & \dots & \dots & \dots & y[k+j-1] \\ u[k+1] & u[k+2] & \dots & \dots & \dots & u[k+j] \\ y[k+1] & y[k+2] & \dots & \dots & \dots & y[k+j] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u[k+i-1] & u[k+i] & \dots & \dots & \dots & u[k+j+i-2] \\ y[k+i-1] & y[k+i] & \dots & \dots & \dots & y[k+j+i-2] \end{bmatrix} \quad (9)$$

$$H_2 = \begin{bmatrix} u[k+i] & u[k+i+1] & \dots & \dots & \dots & u[k+i+j-1] \\ y[k+i] & y[k+i+1] & \dots & \dots & \dots & y[k+i+j-1] \\ u[k+i+1] & u[k+i+2] & \dots & \dots & \dots & u[k+i+j] \\ y[k+i+1] & y[k+i+2] & \dots & \dots & \dots & y[k+i+j] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u[k+2i-1] & u[k+2i] & \dots & \dots & \dots & u[k+2i+j-2] \\ y[k+2i-1] & y[k+2i] & \dots & \dots & \dots & y[k+2i+j-2] \end{bmatrix} \quad (10)$$

Notice that theorem 3 remains valid ! We also introduce the following notations :
$M(p:q,r:s)$ is the submatrix of M at the intersection of rows $p,p+1,\dots,q$ and columns $r,r+1,\dots,s$
$M(:,r:s)$ is the submatrix of M containing columns $r,r+1,\dots,s$
$M(p:q,:)$ is the submatrix of M containing rows $p,p+1,\dots,q$
As an example :

$$H_1 = H(1:mi+li,:)$$

Now let the SVD of $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ be

$$H = U.S.V^t$$

Theorem 5 *The system matrices can be identified from the following set of linear equations*

$$\begin{bmatrix} U_q^t.U_{12}^t.U(m+l+1:(i+1)(m+l),:).S \\ U(mi+li+m+1:(m+l)(i+1),:).S \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} U_q^t.U_{12}^t.U(1:mi+li,:).S \\ U(mi+li+1:mi+li+m,:).S \end{bmatrix}$$

(see section 3 for a definition of U_q and U_{12})

Proof
From section 3 it follows that

$$[x[k+i] \dots x[k+i+j-1]]$$

$$
\begin{aligned}
&= U_q^t.U_{12}^t.H_1 \\
&= U_q^t.U_{12}^t.H(1:mi+li,:) \\
&= U_q^t.U_{12}^t.U(1:mi+li,:).S.V^t
\end{aligned} \tag{11}
$$

Making use of the time-invariance and the block Hankel structure of matrix H, one can easily prove that

$$
\begin{aligned}
&[x[k+i+1] \ \dots \ x[k+i+j]] \\
&\quad = U_q^t.U_{12}^t.H(m+l+1:(i+1)(m+l),:) \\
&\quad = U_q^t.U_{12}^t.U(m+l+1:(i+1)(m+l),:).S.V^t
\end{aligned} \tag{12}
$$

Also, from the definition of H, it follows that

$$
\begin{aligned}
&[u[k+i] \ \dots \ u[k+i+j-1]] \\
&\quad = H(mi+li+1:mi+li+m,:) \\
&\quad = U(mi+li+1:mi+li+m,:).S.V^t
\end{aligned} \tag{13}
$$

and

$$
\begin{aligned}
&[y[k+i] \ \dots \ y[k+i+j-1]] \\
&\quad = H(mi+li+m+1:(m+l)(i+1),:) \\
&\quad = U(mi+li+m+1:(m+l)(i+1),:).S.V^t
\end{aligned} \tag{14}
$$

When equations (11),(12),(13) and (14) are substituted into the following (overdetermined) set of linear equations :

$$
\begin{bmatrix} x[k+i+1] & \dots & x[k+i+j] \\ y[k+i] & \dots & y[k+i+j-1] \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} x[k+i] & \dots & x[k+i+j-1] \\ u[k+i] & \dots & u[k+i+j-1] \end{bmatrix}
$$

one obtains :

$$
\begin{aligned}
&\begin{bmatrix} U_q^t.U_{12}^t.U(m+l+1:(i+1)(m+l),:).S.V^t \\ U(mi+li+m+1:(m+l)(i+1),:).S.V^t \end{bmatrix} \\
&= \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} U_q^t.U_{12}^t.U(1:mi+li,:).S.V^t \\ U(mi+li+1:mi+li+m,:).S.V^t \end{bmatrix}
\end{aligned}
$$

The common (orthogonal) factor V^t can be discarded, thus effectively reducing the number of equations (remember $j \gg \max(mi, li)$),without altering the least squares solution :

$$
\begin{aligned}
&\begin{bmatrix} U_q^t.U_{12}^t.U(m+l+1:(i+1)(m+l),:).S \\ U(mi+li+m+1:(m+l)(i+1),:).S \end{bmatrix} \\
&= \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} U_q^t.U_{12}^t.U(1:mi+li,:).S \\ U(mi+li+1:mi+li+m,:).S \end{bmatrix}
\end{aligned}
$$

$\square$

Note : The common factor S imposes weights on the different equations. Discarding it would alter the least squares solution.

5 Off-line algorithm

The results of the previous sections are summarized into the following off-line algorithm :

Algorithm
Let H be the concatenation of H_1, H_2, defined by equations (9) and (10). The system matrices are then obtained as follows :
1. calculate U and S in the SVD of H

$$H = U.S.V^t = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} . \begin{bmatrix} S_{11} & 0 \\ 0 & 0 \end{bmatrix} .V^t$$

2. calculate the SVD of $U_{12}^t.U_{11}.S_{11}$

$$U_{12}^t.U_{11}.S_{11} = \begin{bmatrix} U_q & U_q^\perp \end{bmatrix} \begin{bmatrix} S_q & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_q^t \\ V_q^{\perp t} \end{bmatrix}$$

3. solve the following set of linear equations

$$\begin{bmatrix} U_q^t.U_{12}^t.U(m+l+1:(i+1)(m+l),:).S \\ U(mi+li+m+1:(m+l)(i+1),:).S \end{bmatrix}$$
$$= \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} U_q^t.U_{12}^t.U(1:mi+li,:).S \\ U(mi+li+1:mi+li+m,:).S \end{bmatrix}$$

It is worth noticing that the system matrices are ultimately identified from U and S only ($H = U.S.V^t$), and that the much larger and much more noise sensitive matrix V is fortunately never used. Even the state vector sequence X_2 does not need to be constructed explicitely.
Besides, this reduction will turn out to be very useful when an adaptive identification algorithm is constructed (see section 6).

Example
The performance of the algorithm has been evaluated on both simulated and industrial data sets. The following example is due to Prof. R.Guidorzi (University of Bologna) (Guidorzi and Rossi 1974). The I/O-sequence was obtained under normal operating conditions of a 120 MW power plant (Pont sur Sambre - France), a system with 5 inputs and 3 outputs. The identified models (for different system order estimates) were evaluated by comparing original and simulated outputs, using the original input signals and the identified model (Figure 1). These simulations demonstrate the remarkable robustness of the identification scheme with respect to over- and underestimation of the system order.

Figure 1: Identification of a power plant : original and reconstructed outputs for different system order estimates.

6 On-line algorithm

The above algorithm is easily converted into an adaptive one, where model updating should account for time-variance. Every time step a new input-output measurement becomes available, defining a new column to be added to the matrix H. On the other hand older measurements should be discarded by successively deleting columns from H. The off-line algorithm of the previous section is then applied to the updated H-matrix.
Instead of using this moving window technique, one can also apply exponential weighting. New columns are still added to H, but instead of deleting columns, all columns are multiplied by a weighting factor $\alpha(\alpha \leq 1)$. This way, a column that was added q time steps earlier, is weighted with a factor α^q, thus effectively reducing the contribution of older data.
Since only U and S in the SVD of H are needed (see section 5), H does not need to be constructed explicitely, since the weighting can be applied to S as well.

Algorithm
Initialize $U_0 = I_{(2mi+2li)\times(2mi+2li)}, S_0 = 0_{(2mi+2li)\times(2mi+2li)}$, m and l being the number of inputs and outputs respectively, $2i$ being the number of block rows in the fictitious matrix H

for $k = 1, \ldots$
1. construct new column *column* to be added to H, using the $2i$ latest I/O-measurements
2. calculate SVD

$$U_k.S_k.V_k^t = [\alpha.U_{k-1}.S_{k-1} \;\; column]$$

and partition

$$U_k.S_k = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} . \begin{bmatrix} S_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

3. calculate the SVD of $U_{12}^t.U_{11}.S_{11}$

$$U_{12}^t.U_{11}.S_{11} = \begin{bmatrix} U_q & U_q^{\perp} \end{bmatrix} \begin{bmatrix} S_q & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_q^t \\ V_q^{\perp t} \end{bmatrix}$$

4. solve the following set of linear equations

$$\begin{aligned} & \begin{bmatrix} U_q^t.U_{12}^t.U(m+l+1:(i+1)(m+l),:).S \\ U(mi+li+m+1:(m+l)(i+1),:).S \end{bmatrix} \\ = & \quad \begin{bmatrix} A & B \\ C & D \end{bmatrix} . \begin{bmatrix} U_q^t.U_{12}^t.U(1:mi+li,:).S \\ U(mi+li+1:mi+li+m,:).S \end{bmatrix} \end{aligned}$$

end

Example
As an example, a second order time-variant system with two inputs and two outputs and sinusoidally varying system poles was identified. Figure (2) shows the identified system poles when the weighting factor is set equal to $1-2^{-4}$.

Figure 2: Identified poles for a second order time-varying system with sinusoidally varying system poles

7 Conclusion

A novel strategy for state space identification from (noisy) I/O-measurements was presented. The system matrices are identified by only applying numerically stable SVD-techniques to a block Hankel matrix (number of columns $\gg$ number of rows), constructed with I/O-data. As it turns out that only the left singular basis is required, both the computational load and the noise sensitivity are considerably reduced. Moreover, the algorithm is easily converted into an adaptive version for slowly time-varying systems, making use of adaptive SVD-algorithms. Extensive simulations have demonstrated the remarkable robustness of the identification scheme with respect to noise and over- and underestimation of the system order.

References

[1] AKAIKE H.,1974. Stochastic theory of minimal realization. IEEE Trans. on Automatic Control, Vol.AC-19, No.6 (pp 667-674).

[2] AKAIKE H.,1975. Markovian representations of stochastic processes by canonical variables. SIAM J. Control, Vol.13 (pp 162-173).

[3] BARAM Y.,1981. Realization and reduction of Markovian models from nonstationary data. IEEE Trans. on Automatic Control, Vol.AC-26, No.6 (pp 1225-1231).

[4] DE MOOR B., MOONEN M., VANDENBERGHE L., VANDEWALLE J.,1987. Identification of linear state space models with singular value decomposition using canonical correlation analysis. Proc. of the International Workshop on SVD and Signal Processing. Grenoble, France. North Holland, Elseviers Science Publishers B.V. (in press)

[5] GOLUB G.H. and VAN LOAN C.F.,1983. Matrix computations. North Oxford Academic Publishing Co., Johns Hopkins University Press.

[6] GUIDORZI R. and ROSSI R.,1974. Identification of a power plant from normal operating records. Automatic Control Theory and Applications, vol.2, No.3 (pp.63-67).

[7] HO B.L. and KALMAN R.E.,1965. Effective construction of linear state variable models from input output data. Proc. 3rd Allerton Conf. (pp.449-459).

[8] HOTELLING H.,1936. Relations between two sets of variates. Biometrika, vol.28 (pp.321-377).

[9] KAILATH T.,1974. A view of three decades of linear filter theory. IEEE Trans. Info. Theory, vol.20 (pp.146-181).

[10] KAILATH T.,1980. Linear system theory. Prentice-Hall, New Jersey.

[11] KALMAN R.E.,1982. System identification from noisy data. Dynamical Systems II, A.R.Bednarek and L.Cesari editors, Academic Press.

[12] KLEMA V.C. and LAUB A.J.,1980. The singular value decomposition: its computation and some applications. IEEE Trans. Aut. Control, vol.AC-25.

[13] KUNG S.Y.,1978. A new identification and model reduction algorithm via singular value decomposition. Proc. 12th Asilomar Conf. on Circuits, Systems and Computers. Pacific Grove. (pp.705-714).

[14] LARIMORE W.E.,1984. System identification, reduced-order filtering and modelling via canonical variate analysis. Proc. of the 1983 Automatic Control Conference. San Francisco.

[15] LAWSON C. and HANSON R.,1974. Solving least squares problems. Prentice Hall Series in Automatic Computation, Englewood Cliffs.

[16] LJUNG L. and SODERSTROM T.,1983. Theory and practice of recursive identification. MIT Press, Cambridge, Massachusetts.

[17] RAMOS J.A. and VERRIEST E.I.,1984. A unifying tool for comparing stochastic realization algorithms and model reduction techniques. Proc. of the 1984 Automatic Control Conference. San Diego.

[18] ZEIGER H.P. and Mc EWEN A.J.,1974. Approximate linear realizations of given dimensions via Ho's algorithm. IEEE Trans. Aut. Control, vol AC-19, (pp 153).