System Identification
Lecture 11

Realization Theory and Subspace Identification

Contents

Realization algorithms and subspace identification

Objective

- Identifying MIMO models from data
- Specifically directed towards state-space (SS) models
- Strong focus on algorithmic approach

Stepwise presentation:

- Realization theory:
  - from Markov parameters (pulse responses) to SS models
  - First for noise-free data; then noisy data
- Subspace identification:
  SS model identification from general i/o data

Realization algorithms

The problem that is addressed in this section is:
given the pulse response (Markov parameters)

$$\{g(t)\}_{t=0,\cdots,\infty}$$

of a finite-dimensional LTI discrete-time dynamical system

$$G(z) = \sum_{t=0}^{\infty} g(t) z^{-t},$$

construct a minimal state space model of the system, in the form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t); \qquad x(0); \\ y(t) &= Cx(t) + Du(t). \end{aligned}$$

such that $g(t) = \begin{cases} CA^{t-1}B & t \geq 1 \\ D & t = 0 \end{cases}$

$$
\mathsf{H}_{n_r,n_c}(G) = \begin{bmatrix} g(1) & g(2) & \cdots & \cdots & g(n_c) \\ g(2) & g(3) & g(4) & \cdots & \vdots \\ g(3) & g(4) & g(5) & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g(n_r) & \cdots & \cdots & \cdots & g(n_r+n_c-1) \end{bmatrix},
$$

$$
= \Gamma_{o,n_r} \cdot \Gamma_{c,n_c}
$$

with

$$
\Gamma_{o,n_r} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n_r-1} \end{bmatrix} \quad \text{observability matrix}
$$

$$
\Gamma_{c,n_c} = \begin{bmatrix} B & AB & \cdots & A^{n_c-1}B \end{bmatrix} \quad \text{controllability matrix.}
$$

For sufficiently large n_r, n_c the rank of the Hankel matrix satisfies

$$rank\ \mathrm{H}_{n_r,n_c}(G) = n$$

where n is the McMillan degree of $G(z)$,
i.e. the minimal state dimension of a state-space model for G

(see also Model Reduction course)

When taking the Markov parameter sequence as a starting point: When given a sufficently large Hankel matrix with Markov parameters, then for any (full rank) matrix decomposition

$$\mathrm{H}_{n_r,n_c}(G) = H_1 \cdot H_2$$

with $H_1 \in \mathbb{R}^{n_r p \times n}$ and $H_2 \in \mathbb{R}^{n \times n_c m}$ satisfying

$$rank\ H_1 = rank\ H_2 = rank\ \mathrm{H}_{n_r,n_c}(G) = n$$

there exist matrices A, B, C from an n-dimensional state space model, such that

$$\begin{aligned} H_1 &= \Gamma_{o,n_r} \quad \text{and} \\ H_2 &= \Gamma_{c,n_c}. \end{aligned}$$

These matrices (A, B, C) can be constructed as follows:

- When given $H_1 = \Gamma_{o,n_r}$, i.e.

$$H_1 = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n_r-1} \end{bmatrix},$$

  the matrix C can be extracted by taking the first p rows of the matrix;

- When given $H_2 = \Gamma_{o,n_r}$, i.e.

$$H_2 = \begin{bmatrix} B & AB & \cdots & A^{n_c-1}B \end{bmatrix},$$

  the matrix B can be extracted by taking the first m columns of the matrix.

- The matrix A can be isolated by using the shifted Hankel matrix:

$$\overleftarrow{\mathrm{H}}_{n_r,n_c}(G) := \begin{bmatrix} g(2) & g(3) & \cdots & \cdots & g(n_c+1) \\ g(3) & g(4) & g(5) & \cdots & \vdots \\ g(4) & g(5) & g(6) & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g(n_r+1) & \cdots & \cdots & \cdots & g(n_r+n_c) \end{bmatrix},$$

by using

$$\overleftarrow{\mathrm{H}} = H_1 \cdot A \cdot H_2$$

it follows that

$$H_1^+ \cdot \overleftarrow{\mathrm{H}} \cdot H_2^+ = A$$

for pseudo inverses H_1^+, H_2^+, satisfying $H_1^+ H_1 = H_2 H_2^+ = I_n$.

Calculations can be done with Singular Value Decomposition (SVD):

$$\mathrm{H} = U_n \Sigma_n V_n^T \qquad U_n^T U_n = V_n^T V_n = I$$

then choosing

$$\begin{aligned} H_1 &= U_n \Sigma_n^{\frac{1}{2}} \\ H_2 &= \Sigma_n^{\frac{1}{2}} V_n^T \end{aligned}$$

directly leads to

$$\begin{aligned} H_1^+ &= \Sigma_n^{-\frac{1}{2}} U_n^T \\ H_2^+ &= V_n \Sigma_n^{-\frac{1}{2}} \end{aligned}$$

and

$$A = \Sigma_n^{-\frac{1}{2}} U_n^T \cdot \overleftarrow{\mathrm{H}} \cdot V_n \Sigma_n^{-\frac{1}{2}}.$$

Algorithm of Ho & Kalman (1966)

- Construct H with row and column dimension larger than $n = rank\ \mathrm{H}$
- Apply SVD: $\mathrm{H} = U_n \Sigma_n V_n^T$
- Construct C equal to the first block row of $U_n \Sigma_n^{1/2}$
- Construct B equal to the first block column of $\Sigma_n^{1/2} V_n^T$
- Construct the shifted Hankel matrix $\overleftarrow{\mathrm{H}}$
- Determine the matrix A according to

$$\overleftarrow{\mathrm{H}} = U_n \Sigma_n^{1/2} \cdot A \cdot \Sigma_n^{1/2} V_n^T$$

$$A = \Sigma_n^{-1/2} U_n^T \cdot \overleftarrow{\mathrm{H}} \cdot V_n \Sigma_n^{-1/2}$$

The resulting model (A, B, C) satisfies

$$g(t) = CA^{t-1}B \quad t \geq 1$$

Example

$g(0), g(1) \cdots$ is given by

$$0, 1, 0.5, 0.25, 0.125, 0.0625, 0.03125, \cdots$$

$D = 0.$

$$\mathrm{H}_{3,3} = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 0.25 & 0.125 \\ 0.25 & 0.125 & 0.0625 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \\ 0.25 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0.5 & 0.25 \end{bmatrix}$$

$B = 1; C = 1.$

$$\overleftarrow{\mathrm{H}}_{3,3} = \begin{bmatrix} 0.5 & 0.25 & 0.125 \\ 0.25 & 0.125 & 0.0625 \\ 0.125 & 0.0625 & 0.03125 \end{bmatrix}$$

A is obtained through:

$$\overleftarrow{H}_{3,3} = \begin{bmatrix} 1 \\ 0.5 \\ 0.25 \end{bmatrix} \cdot A \cdot [1 \quad 0.5 \quad 0.25]$$

or

$$[1\ 0\ 0]\ \overleftarrow{H}_{3,3} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = A$$

$$\begin{aligned} A &= [1\ 0\ 0] \begin{bmatrix} 0.5 & 0.25 & 0.125 \\ 0.25 & 0.125 & 0.0625 \\ 0.125 & 0.0625 & 0.03125 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= 0.5 \end{aligned}$$

- Although the algorithm is developed for exact minimal realization (noise-free data), it is also applied to "estimate" a low-order model on the basis of noisy Markov parameters.
- In the SVD, the smaller singular values are then simply set to 0, to achieve a model of order (state space dimension) n, i.e. the number of singular values $\neq 0$.

Approximate realization for noisy data

Observation:
For noisy data: H has rank equal to its dimension $\rightarrow$ larger H leads to higher order models

Reasoning:

(1) Approximate H by a matrix with reduced rank n

(2) Apply the Ho/Kalman algorithm to this matrix

Resulting state space model has state dimension n

Reduced rank approximation of a matrix

Given a matrix P with SVD:

$$P = [U_k \mid \bar{U}] \begin{bmatrix} \Sigma_k & 0 \\ 0 & \bar{\Sigma} \end{bmatrix} [V_k \mid \bar{V}]^T$$

and define:

$$P_k = U_k \Sigma_k V_k^T$$

Then it holds that P_k minimizes

$$\|P - \tilde{P}\|_2 \quad \text{en} \quad \|P - \tilde{P}\|_F$$

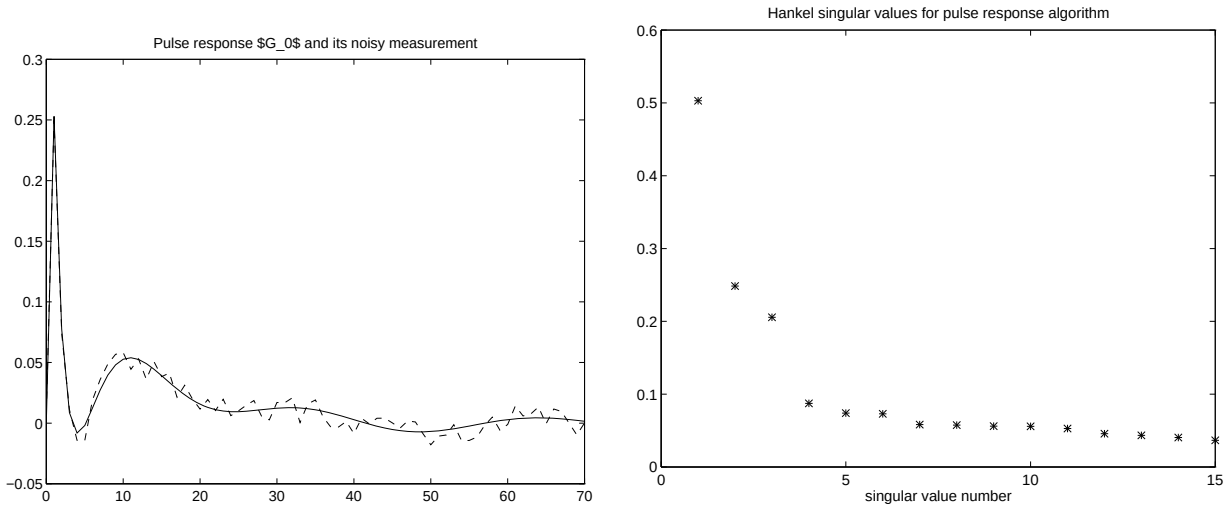
over all matrices $\tilde{P}$ of rank k.
In other words: P_k is the best rank-k approximation of P in the sense of these norms.

Singular values σ_i provide information on approximation:

$$\|P - P_k\|_2 = \sigma_{k+1}$$

$$\|P - P_k\|_F = \left(\sum_{i=k+1}^{min(q,r)} \sigma_i^2 \right)^{\frac{1}{2}}$$

Singular value plot allows for order selection:



Left: Exact (solid) and noisy (dashed) pulse response of G_0; Right: singular values of $H_{35,35}$

Algorithm of Kung:

- Construct H and apply SVD: $\mathrm{H} = U\Sigma V^T$
- Evaluate singular values and choose a reduced rank n
- Construct approximate matrix $\mathrm{H}(n) = U_n\Sigma_n V_n^T$
- Choose C as the first (block) row of $U_n\Sigma_n^{1/2}$
- Choose B as the first (block) column of $\Sigma_n^{1/2}V_n^T$
- Construct A according to

$$\Sigma_n^{-1/2}U_n^T \cdot \overleftarrow{\mathrm{H}} \cdot V_n\Sigma_n^{-1/2}$$

with $\overleftarrow{\mathrm{H}}$ the (block) shifted matrix of original Markov parameteres.

Resulting model (A, B, C) has state space dimension n

Summary

- Realization theory provides attractive algorithms for estimating state-space models on the basis of impulse response / Markov parameter data
- Simple tools from linear algebra
- Well suited for MIMO models
- Algorithms have intrinsic option for model order selection (through SVD)

Subspace identification - intro

- Alternative identification framework
- Developed in the spirit of realization theory
- Estimation of state space models
- Well suited for multivariable systems
- Incorporated order selection mechanism
- Strongly algorithm driven (less emphasis on analysis)

Content

- State space models
- Identification problem
- Step 1: (C, A) known: Linear regression
- Step 2: Observability matrix estimation
- Summary

Presentation follows the reasoning in Ljung (1999), section 10.6

State space models

Discrete-time LTI system: $y(t) = G_0(q)u(t) + v(t)$
can be written in state-space form:

$$\begin{aligned} x(t+1) &= A_0 x(t) + B_0 u(t) \\ y(t) &= C_0 x(t) + D_0 u(t) + v(t) \end{aligned}$$

where $G_0(z) = C_0(zI - A_0)^{-1}B_0 + D_0$.

Model is not unique: for any nonsingular square matrix T, and $\bar{x}(t) := Tx(t)$ it follows that

$$\begin{aligned} \bar{x}(t+1) &= \bar{A}_0 \bar{x}(t) + \bar{B}_0 u(t) \\ y(t) &= \bar{C}_0 \bar{x}(t) + D_0 u(t) + v(t) \end{aligned}$$

with $\bar{A}_0 = TA_0T^{-1}$, $\bar{B}_0 = TB_0$, $\bar{C}_0 = C_0T^{-1}$.

If v is colored noise $v(t) = H_0(q)e(t)$ this can be represented in the state-space model by

$$\begin{aligned} x(t+1) &= A_0 x(t) + B_0 u(t) + K_0 e(t) \\ y(t) &= C_0 x(t) + D_0 u(t) + e(t) \end{aligned}$$

met $\mathrm{E}[ee^T] = \lambda I$. For Output error (OE) structure: $K_0 = 0$.

Most general form:

$$\begin{aligned} x(t+1) &= A_0 x(t) + B_0 u(t) + w(t) \\ y(t) &= C_0 x(t) + D_0 u(t) + \nu(t) \end{aligned}$$

with w, ν white noise processes, with $\mathrm{E}[\nu\nu^T] = R$, $\mathrm{E}[ww^T] = Q$, $\mathrm{E}[w\nu^T] = S$.

Identification problem

Consider data generating system:

$$\begin{aligned} x(t+1) &= A_o x(t) + B_o u(t) + w(t) \\ y(t) &= C_o x(t) + D_o u(t) + \nu(t) \end{aligned}$$

with

$$\mathrm{E}\left[\begin{pmatrix} w \\ \nu \end{pmatrix}\begin{pmatrix} w^T & \nu^T \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}.$$

Determine of the basis of $\{u(t), y(t)\}_{t=1,\ldots,N}$:

- The order n (dimension of x) of the system
- Matrices $[A, B, C, D]$, modulo a similarity transformation T
- The covariance matrices $[Q, R, S]$ (optional)

Focus on consistency. No particular identification criterion.

Subspace identification through observability matrix estimation

$$G(q) = C(qI - A)^{-1}B + D$$

Stepwise procedure:

1. If (C, A) would be known then the identification problem of (B, D) is relatively simple (linear regression type of problem)
2. (C, A) can be estimated through estimating the observability matrix:

$$O_r = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix}$$

Step 1: Estimating B, D when A, C are known

Use the predictor:

$$\hat{y}(t|t-1;\theta) := [C(qI - A)^{-1}B(\theta) + D(\theta)]u(t)$$

with entries of B and D parametrized.
This predictor is linear-in-the-parameters, and so can be written as

$$\hat{y}(t|t-1;\theta) = \varphi^T(t)\theta$$

where vector θ is composed of the entries of B and D:

- θ is a $(n+p)m \times 1$ vector.
- $\varphi^T(t)$ is a matrix of dimensions $p \times (n+p)m$,
- $\varphi(t)$ contains past (and present) values of $u(t)$.

Step 1: Estimating B, D when A, C are known (cont'd)

A least squares solution for θ can then simply be obtained:

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} [y(t) - \varphi^T(t)\theta]^T \Lambda^{-1} [y(t) - \varphi^T(t)\theta]$$

through a convex optimization problem, with an analytical result (for $\Lambda = I$):

$$\hat{\theta}_N = [\Phi^T \Phi]^{-1} \Phi^T \; Y$$

with

$$Y := \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \qquad \Phi := \begin{bmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{bmatrix}$$

Step 2: Estimating C, A through estimating observability matrix

We start with the output equation:

$$y(t) = Cx(t) + Du(t) + \nu(t).$$

Writing this down for $t+1$ (assuming for simplicity that $w(t) \equiv 0$):

$$\begin{aligned} y(t+1) &= Cx(t+1) + Du(t+1) + \nu(t+1) \\ &= CAx(t) + Du(t+1) + CBu(t) + \nu(t+1) \end{aligned}$$

and for $t+2$:

$$\begin{aligned} y(t+2) &= CAx(t+1) + Du(t+2) + \nu(t+2) \\ &= CA^2x(t) + Du(t+2) + CBu(t+1) + CABu(t) + \nu(t+2) \end{aligned}$$

Continuing this, and denoting

$$Y_r(t) = \begin{bmatrix} y(t) \\ \vdots \\ y(t+r-1) \end{bmatrix}, \; U_r(t) = \begin{bmatrix} u(t) \\ \vdots \\ u(t+r-1) \end{bmatrix}, V(t) = \begin{bmatrix} \nu(t) \\ \vdots \\ \nu(t+r-1) \end{bmatrix}$$

leads to

$$Y_r(t) = O_r x(t) + S_r U_r(t) + V(t)$$

with extended observability matrix

$$O_r = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix} \quad \text{and} \quad S_r = \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ CA^r B & \cdots & \cdots & \cdots & D \end{bmatrix}$$

With $\mathbf{Y} := [Y_r(1) \ \cdots \ Y_r(N)]$, and similar notation for x, U_r, V this becomes

$$\mathbf{Y} = O_r \mathbf{X} + S_r \mathbf{U} + \mathbf{V}. \qquad \mathbf{Y} \in \mathbb{R}^{r \times N}$$

Estimation of O_r can now be done through projections/correlations.

$$\mathbf{Y} = O_r \mathbf{X} + S_r \mathbf{U} + \mathbf{V}.$$

I.e. Process the equation with right multiplications such that the terms with $\mathbf{U}$ and $\mathbf{V}$ disappear.

Removal of input-dependent term

$$\mathbf{Y} = O_r \mathbf{X} + S_r \mathbf{U} + \mathbf{V}.$$

Right multiplication with projector matrix $\mathbf{U}^\perp := I - \mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^{-1}\mathbf{U}$ delivers

$$\mathbf{Y}\mathbf{U}^\perp = O_r \mathbf{X}\mathbf{U}^\perp + \mathbf{V}\mathbf{U}^\perp$$

since

$$\mathbf{U}\mathbf{U}^\perp = \mathbf{U}[I - \mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^{-1}\mathbf{U}] = \mathbf{U} - \mathbf{U} = 0$$

$\mathbf{U}^\perp$ provides projection onto the space that is orthogonal to (the row space of) U.

Removal of noise-dependent term

$$\mathbf{Y}\mathbf{U}^{\perp} = O_r\mathbf{X}\mathbf{U}^{\perp} + \mathbf{V}\mathbf{U}^{\perp}$$

Design an instrument matrix

$$\mathbf{\Phi} = [\phi(1) \ \cdots \ \phi(N)] \ \text{ with } \phi(t) \in \mathbb{R}^s, s \geq n$$

leading to

$$\frac{1}{N}\mathbf{Y}\mathbf{U}^{\perp}\mathbf{\Phi}^T = O_r\frac{1}{N}\mathbf{X}\mathbf{U}^{\perp}\mathbf{\Phi}^T + \frac{1}{N}\mathbf{V}\mathbf{U}^{\perp}\mathbf{\Phi}^T$$

such that

- $\lim_{N\to\infty} \frac{1}{N}\mathbf{V}\mathbf{U}^{\perp}\mathbf{\Phi}^T = 0$, and
- $\lim_{N\to\infty} \frac{1}{N}\mathbf{X}\mathbf{U}^{\perp}\mathbf{\Phi}^T = \tilde{T}$, having full row rank

$$\underbrace{\frac{1}{N}\mathbf{Y}\mathbf{U}^{\perp}\mathbf{\Phi}^{T}}_{G_N} = O_r \underbrace{\frac{1}{N}\mathbf{X}\mathbf{U}^{\perp}\mathbf{\Phi}^{T}}_{\tilde{\mathbf{T}}_\mathbf{N}} + \frac{1}{N}\mathbf{V}\mathbf{U}^{\perp}\mathbf{\Phi}^{T}$$

$$\Downarrow N \to \infty$$

$$G = O_r \tilde{\mathbf{T}}$$

Then G_N is a noisy estimate of $O_r \tilde{T}$, and for N sufficiently large, it holds that

The column space of $G_N \equiv$ The column space of O_r

Column space of G_N is determined through svd: $G_N = U\Sigma V^T$.
C and A can be determined from the column space of O_r, by setting

$$O_r = U$$

So the sequence of steps becomes:

- Calculate G_N and determine $G_N = U\Sigma V^T$;
- Set $U = O_r$, i.e.

$$U = \begin{bmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{r-1} \end{bmatrix} = O_r$$

- Set C equal to the first p rows of O_r
- Construct O_{r-1} and construct its (block-)shifted version $O_{r-1}^{\uparrow}$
- Determine A on the basis of $O_{r-1}^{\uparrow} = O_{r-1}A$, i.e.

$$A = O_{r-1}^{+} O_{r-1}^{\uparrow}$$

On the basis of the SVD:

$$G_N = U\Sigma V^T$$

an approximation

$$G_N = U_n\Sigma_n V_n^T$$

can be made, to limit the column space to dimension n.

This will lead to a state space model with state dimension n.

The SVD of G_N acts as a model order selection mechanism

Choice of instruments

The two conditions:

- $\lim_{N\to\infty} \frac{1}{N}\mathbf{V}\mathbf{U}^{\perp}\mathbf{\Phi}^{T} = 0$, and
- $\lim_{N\to\infty} \frac{1}{N}\mathbf{X}\mathbf{U}^{\perp}\mathbf{\Phi}^{T} = \tilde{T}$, having full row rank

have to be satisfied.

With $\mathbf{U}^{\perp} = I - \mathbf{U}^{T}(\mathbf{U}\mathbf{U}^{T})^{-1}\mathbf{U}$, and u being uncorrelated with ν (in open-loop case), the crucial condition to satisfy is

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{V}\mathbf{\Phi}^{T} = 0$$

$\lim_{N\to\infty} \frac{1}{N}\mathbf{V}\mathbf{\Phi}^T = 0$ implies

$$\lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} \nu(1) & \cdots & \nu(N) \\ \nu(2) & \cdots & \nu(N+1) \\ \vdots & \vdots & \vdots \\ \nu(r) & \cdots & \nu(r+N-1) \end{bmatrix} \begin{bmatrix} \phi^T(1) \\ \phi^T(2) \\ \vdots \\ \phi^T(N) \end{bmatrix} = 0$$

which through the whiteness of $\nu(t)$ can be achieved (asymptotically) if

$$\phi(t) \text{ is uncorrelated with } \nu(t).$$

e.g. (in open loop when u and ν are independent):

$$\phi(t) = [y(t-1) \ \ldots \ y(t-s_1) \ u(t-1) \ \ldots \ u(t-s_2)]^T$$

Subspace ID through observability - algorithm

- Choose $r \geq n$ and instrument $\phi(t)$, and construct G_N
- Construct a low rank approximation

$$W_1 G_N W_2 = U \Sigma V^T \approx U_n \Sigma_n V_n^T$$

  with user-chosen weighting matrices W_1, W_2.
- Construct $\hat{O}_r = W_1^{-1} U_1 T$ with T any nonsingular matrix
- Determine C, A from $\hat{O}_r$
- With C, A given, determine B, D through an LS-algorithm.

Weighting matrices W_1, W_2 are chosen differently for different methods, as N4SID (Van Overschee et al.), MOESP (Verhaegen), CVA (Larimore), IVM (Viberg).

Summary subspace identification

- Identification methods based on LS algorithms and numerical tools as svd (projections)
- No optimization, numerical robust and simply applicable also to MIMO systems
- Because of multi-step algorithms, approximation properties ($\mathcal{S} \notin \mathcal{M}$) are hard to analyze
- Attractive to use, also as initial model for PE/ML optimization
- Extensions available for closed-loop and f-domain data

Literature

- W.E. Larimore (1990). Canonical variate analysis in identification, filtering and adaptive control. *Proc. 29th IEEE Conf. Decision and Control*, Honolulu, HI, pp. 596-604.
- P. Van Overschee and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic and stochastic systems. *Automatica*, Vol. 30, no. 1, pp. 75-93.
- P. Van Overschee and B. De Moor (1996). *Subspace Identification for Linear Systems; Theory, Implementation,m Applications*. Kluwer Aced. Publ., Dordrecht.
- M. Verhaegen and P. Dewilde (1992). Subspace model identification part 1. The output-error state-space model identification class of algorithms. *Int. J. Control*, Vol. 56, no. 5, pp. 1187-2010.
- M. Verhaegen and V. Verdult (2007). *Filtering and System Identification - A Least Squares Approach*. Cambridge University Press, UK.
- M. Viberg (1995). Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, Vol. 31, no. 12, pp. 1835-1851.