System identification in the behavioral setting

Ivan Markovsky

K.U.Leuven, ESAT-SISTA

Joint work with

J. C. Willems, P. Rapisarda, B. De Moor, and S. Van Huffel

Outline

- Exact identification
- Algorithms for exact identification
- Relation to deterministic subspace identification
- Approximate identification
- Conclusions

Exact identification

An exact identification problem

Problem P1 (Exact identification)

Given two vector time series

$$u_{\mathrm{d}} = \big(u_{\mathrm{d}}(1), \ldots, u_{\mathrm{d}}(T)\big) \in (\mathbb{R}^{\mathtt{m}})^T \qquad \text{"inputs"}$$

$$y_{\mathrm{d}} = \big(y_{\mathrm{d}}(1), \ldots, y_{\mathrm{d}}(T)\big) \in (\mathbb{R}^{\mathtt{p}})^T \qquad \text{"outputs"}$$

find $\mathtt{n} \in \mathbb{N}$ and an LTI system Σ of minimal order $\mathtt{n}$, with $\mathtt{m}$ inputs and $\mathtt{p}$ outputs, such that $(u_{\mathrm{d}}, y_{\mathrm{d}})$ is a trajectory of Σ.

What does it mean "is a trajectory of"?

What does it mean "is a trajectory of"?

let σ be the shift operator $\sigma x(t) = x(t+1)$ and let Σ be defined by a minimal state space representation

$$\Sigma: \quad \sigma x = Ax + Bu, \quad y = Cx + Du \tag{1}$$

$(u_\mathrm{d}, y_\mathrm{d})$ is a trajectory of Σ if there exists $x_\mathrm{ini} \in \mathbb{R}^\mathtt{n}$, such that

$$y_\mathrm{d} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}_T(A,C)} x_\mathrm{ini} + \begin{bmatrix} D & & & & \\ CB & D & & & \\ CAB & CB & D & & \\ \vdots & \ddots & \ddots & \ddots & \\ CA^{T-1}B & \cdots & CAB & CB & D \end{bmatrix} u_\mathrm{d}$$

i.e., y_d is the response of Σ under input u_d and initial condition x_ini

Comments

- P1 is an exact fitting problem, a most basic system id. problem
- easily generalizable to a set of N time series $u_{\mathrm{d},1}, \dots, u_{\mathrm{d},N} \in (\mathbb{R}^\mathtt{m})^T$ and $y_{\mathrm{d},1}, \dots, y_{\mathrm{d},N} \in (\mathbb{R}^\mathtt{p})^T$
- the realization problem (impulse response $\mapsto (A,B,C,D)$) is a special case of P1 for a set of $\mathtt{m}$ time series
- while $\mathtt{m}$ is given, finding $\mathtt{n}$ is part of the problem

  in fact, any observable system of order $\mathtt{n} \geq \mathtt{p}T$ is a (trivial) solution
- we are actually interested is a solution of a minimal order

An exact identification problem (revised)

Problem P1' (Exact identification)

Given two vector time series

$$u_\mathrm{d} = \big(u_\mathrm{d}(1), \dots, u_\mathrm{d}(T)\big) \in (\mathbb{R}^\mathtt{m})^T \qquad \text{"inputs"}$$
$$y_\mathrm{d} = \big(y_\mathrm{d}(1), \dots, y_\mathrm{d}(T)\big) \in (\mathbb{R}^\mathtt{p})^T \qquad \text{"outputs"}$$

find the smallest $\mathtt{n} \in \mathbb{N}$ and an LTI system Σ of order $\mathtt{n}$, with $\mathtt{m}$ inputs and $\mathtt{p}$ outputs, such that $(u_\mathrm{d}, y_\mathrm{d})$ is a trajectory of Σ.

Behavior and representation of a system

the behavior of an LTI system Σ is the set $\mathscr{B}$ of all trajectories $w := (u,y)$ that Σ can possibly generate

$\mathscr{B}|_{[1,t]}$ — restriction of the behavior to the interval $[1,t]$

a representation of Σ is an equation whose solution set is equal to $\mathscr{B}$

e.g., the inputs/state/output repr. (1), and the difference equation repr.

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_l w(t+l) = 0, \quad \text{for all } t, \quad \text{where } R_i \in \mathbb{R}^{g \times (\mathtt{m}+\mathtt{p})}$$

also called kernel representation because

$$\mathscr{B} = \ker\big(R(\sigma)\big), \qquad \text{where} \quad R(\xi) := \sum_{i=0}^{l} R_i \xi^i$$

Set of LTI systems with a fixed complexity

$\mathscr{L}^{\mathtt{w},\mathtt{n}}_{\mathtt{m},\mathtt{l}}$ — set of all LTI systems with

- $\mathtt{w}$ (external) variables,
- at most $\mathtt{m}$ inputs,
- minimal state dimension at most $\mathtt{n}$, and
- lag (= observability index) at most $\mathtt{l}$

for $t \geq \mathtt{n}$, $\quad \dim(\mathscr{B}|_{[1,t]}) \;\leq\; t\mathtt{m}+\mathtt{n} \;\leq\; t\mathtt{m}+\mathtt{p}\mathtt{l} \qquad (\mathtt{p}(\mathtt{l}-1) \leq \mathtt{n} \leq \mathtt{p}\mathtt{l})$

$\Longrightarrow \quad (\mathtt{m},\mathtt{n})$ and $(\mathtt{m},\mathtt{l})$ specify the complexity

An exact identification problem (revised)

Problem P2 (Exact identification)

Given a vector time series

$$w_{\mathrm{d}} = \big(w_{\mathrm{d}}(1),\dots,w_{\mathrm{d}}(T)\big) \in (\mathbb{R}^{\mathtt{w}})^T$$

find the smallest $\mathtt{m} \in \mathbb{N}$ and $\mathtt{l} \in \mathbb{N}$ and an LTI system $\mathscr{B} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\mathtt{l}}$, such that $w_{\mathrm{d}} \in \mathscr{B}$.

comments:

- no separation between inputs and outputs
- the complexity is defined by $(\mathtt{m},\mathtt{l})$

Most powerful unfalsified model

The most powerful unfalsified model in the model class $\mathscr{L}^{\mathtt{w}}_{\mathtt{m},\mathtt{l}}$ of a time series $w_{\mathrm{d}} \in (\mathbb{R}^{\mathtt{w}})^T$ is the system $\mathscr{B}_{\text{mpum}}$ that is

1. in the model class, *i.e.*, $\mathscr{B}_{\text{mpum}} \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\mathtt{l}}$,
2. unfalsified, *i.e.*, $w_{\mathrm{d}} \in \mathscr{B}_{\text{mpum}}|_{[1,T]}$, and
3. most powerful among all LTI unfalsified systems, *i.e.*,

$$\mathscr{B}' \in \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\mathtt{l}} \text{ and } w_{\mathrm{d}} \in \mathscr{B}'|_{[1,T]} \quad \Longrightarrow \quad \mathscr{B}_{\text{mpum}}|_{[1,T]} \subseteq \mathscr{B}'|_{[1,T]}.$$

the MPUM need not exist, but if it does, then it is unique

Identifiability question

P2 is the problem of computing the MPUM

the following related question is of interest:

Suppose that $w_{\mathrm{d}} \in \mathscr{B} \in \mathscr{L}^{\mathtt{w}}$ and upper bounds $\mathtt{n}_{\max}$ and $\mathtt{l}_{\max}$ of the order $\mathtt{n}$ and the lag $\mathtt{l}$ of $\mathscr{B}$ are given.

$$\boxed{\text{Under what conditions is } \mathscr{B}_{\text{mpum}}(w_{\mathrm{d}}) \text{ equal to the system } \mathscr{B}?}$$

the answer is given by the following lemma

Fundamental Lemma

Let $\mathscr{B} \in \mathscr{L}^{\mathtt{w},\mathtt{n}}_{\mathtt{m}}$ be controllable and let $w_{\mathrm{d}} := (u_{\mathrm{d}}, y_{\mathrm{d}}) \in \mathscr{B}|_{[1,T]}$.
Then, if u_{d} is persistently exciting of order $L + \mathtt{n}$,

$$\text{image}\left(\begin{bmatrix} w_{\mathrm{d}}(1) & w_{\mathrm{d}}(2) & w_{\mathrm{d}}(3) & \cdots & w_{\mathrm{d}}(T-L+1) \\ w_{\mathrm{d}}(2) & w_{\mathrm{d}}(3) & w_{\mathrm{d}}(4) & \cdots & w_{\mathrm{d}}(T-L+2) \\ w_{\mathrm{d}}(3) & w_{\mathrm{d}}(4) & w_{\mathrm{d}}(5) & \cdots & w_{\mathrm{d}}(T-L+3) \\ \vdots & \vdots & \vdots & & \vdots \\ w_{\mathrm{d}}(L) & w_{\mathrm{d}}(L+1) & w_{\mathrm{d}}(L+2) & \cdots & w_{\mathrm{d}}(T) \end{bmatrix} \right) = \mathscr{B}|_{[1,L]}$$

$\Longrightarrow$ under the conditions of the FL, any L samples long response y of $\mathscr{B}$ can be obtained as $y = \mathscr{H}_L(y_{\mathrm{d}})g$, for certain g $\rightsquigarrow$ algorithms

$\Longrightarrow$ with $L = \mathtt{l}_{\max} + 1$, the FL gives conditions for identifiability

Persistency of excitation

the sequence $u_{\mathrm{d}} = \big(u_{\mathrm{d}}(1), \ldots, u_{\mathrm{d}}(T)\big)$ is persistently exciting of order L if the Hankel matrix

$$\mathscr{H}_L(u_{\mathrm{d}}) := \begin{bmatrix} u_{\mathrm{d}}(1) & u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & \cdots & u_{\mathrm{d}}(T-L+1) \\ u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & u_{\mathrm{d}}(4) & \cdots & u_{\mathrm{d}}(T-L+2) \\ u_{\mathrm{d}}(3) & u_{\mathrm{d}}(4) & u_{\mathrm{d}}(5) & \cdots & u_{\mathrm{d}}(T-L+3) \\ \vdots & \vdots & \vdots & & \vdots \\ u_{\mathrm{d}}(L) & u_{\mathrm{d}}(L+1) & u_{\mathrm{d}}(L+2) & \cdots & u_{\mathrm{d}}(T) \end{bmatrix}$$

is of full row rank

Algorithms for exact identification

($w_{\mathrm{d}} \mapsto$ representation of the MPUM)

Overview of algorithms

1. $w_{\mathrm{d}} \mapsto R(\xi)$
2. $w_{\mathrm{d}} \mapsto$ impulse response H
3. $w_{\mathrm{d}} \mapsto (A,B,C,D)$ (possibly balanced)

3.a. $w_{\mathrm{d}} \mapsto R(\xi) \mapsto (A,B,C,D)$ or $w_{\mathrm{d}} \mapsto H \mapsto (A,B,C,D)$

3.b. $w_{\mathrm{d}} \mapsto \mathscr{O}_{\mathtt{l}_{\max}+1}(A,C) \mapsto (A,B,C,D)$

3.c. $w_{\mathrm{d}} \mapsto \big(x_{\mathrm{d}}(1), \ldots, x_{\mathrm{d}}(\mathtt{n}_{\max} + \mathtt{m} + 1)\big) \mapsto (A,B,C,D)$

$w_d \mapsto R(\xi)$

under the assumptions of the FL, $\text{image}\big(\mathscr{H}_{l_{\max}+1}(w_d)\big) = \mathscr{B}|_{[1,l_{\max}+1]}$

$\implies$ a basis for left $\ker\big(\mathscr{H}_{l_{\max}+1}(w_d)\big)$ defines a kernel repr. of $\mathscr{B}$

let

$$\begin{bmatrix} \tilde{R}_0 & \tilde{R}_1 & \cdots & \tilde{R}_{l_{\max}} \end{bmatrix} \mathscr{H}_{l_{\max}+1}(w_d) = 0, \quad \text{where } \tilde{R}_i \in \mathbb{R}^{g \times w}$$

and define $\tilde{R}(\xi) = \sum_{i=0}^{l_{\max}} \xi^i \tilde{R}_i$

then $\mathscr{B} = \ker\big(\tilde{R}(\sigma)\big)$ is, in general, a nonminimal kernel representation

$w_d \mapsto R(\xi)$

$\tilde{R}$ can be made minimal by standard polynomial linear algebra alg.

find a unimodular matrix U, such that

$$U\tilde{R} = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad \text{and } R \text{ is full row rank}$$

then $\ker\big(R(\sigma)\big) = 0$ is minimal

refinements:

- efficient recursive computation (exploiting the Hankel structure)
- as a byproduct find an input/output partition of the variables
- find a shortest lag kernel representation (*i.e.*, R row proper)

$w_d \mapsto H$

under the conditions of FL, there is G, such that $H = \mathscr{H}_t(y_d)G$

the problem reduces to the one of finding a particular G. Define

$$\begin{bmatrix} \mathscr{H}_{l_{\max}+t}(u_d) \\ \mathscr{H}_{l_{\max}+t}(y_d) \end{bmatrix} =: \begin{bmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{bmatrix} \qquad \begin{aligned} \text{row dim}(U_p) &= \text{row dim}(Y_p) = l_{\max} \\ \text{row dim}(U_f) &= \text{row dim}(Y_f) = t \end{aligned}$$

let u_d be p.e. of order $t + l_{\max} + n_{\max}$. Then there is G, such that

$$\begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} G = \begin{bmatrix} 0 \\ 0 \\ \begin{bmatrix} I_m \\ 0 \end{bmatrix} \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{zero ini. conditions} \\ \leftarrow \text{ impulse input} \end{matrix} \tag{2}$$

$$Y_f \quad G = H$$

$w_d \mapsto H$

block algorithm for computation of $\big(H(0), \ldots, H(t-1)\big)$:

1. Input: u_d, y_d, $l_{\max}$, and t.
2. Solve the system of equations (2). Let $\bar{G}$ be the computed solution.
3. Compute $H = Y_f \bar{G}$.
4. Output: the first t samples of the impulse response H.

refinements:

- solve (2) efficiently by exploiting the Hankel structure
- do the computations iteratively for pieces of $H \rightsquigarrow$ iterative algorithm
- automatically choose t, for a sufficient decay of H

$w_d \mapsto (A,B,C,D)$

- $w_d \mapsto H(0:2\mathtt{l}_{\max})$ or $R(\xi) \xrightarrow{\text{realization}} (A,B,C,D)$
- $w_d \mapsto \mathscr{O}_{\mathtt{l}_{\max}+1}(A,C) \xrightarrow{(3)} (A,B,C,D)$
- $w_d \mapsto \big(x_d(1),\dots,x_d(\mathtt{n}_{\max}+\mathtt{m}+1)\big) \xrightarrow{(4)} (A,B,C,D)$

(3) and (4) are easy:

$$\mathscr{O}_{\mathtt{l}_{\max}+1}(A,C) \mapsto (A,C) \quad \text{and} \quad (u_d,y_d,A,C) \mapsto (B,C,x_{\text{ini}}) \tag{3}$$

$$\begin{bmatrix} x_d(2) & \cdots & x_d(\mathtt{n}_{\max}+\mathtt{m}+1) \\ y_d(1) & \cdots & y_d(\mathtt{n}_{\max}+\mathtt{m}) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_d(1) & \cdots & x_d(\mathtt{n}_{\max}+\mathtt{m}) \\ u_d(1) & \cdots & u_d(\mathtt{n}_{\max}+\mathtt{m}) \end{bmatrix} \tag{4}$$

$\mathscr{O}_{\mathtt{l}_{\max}+1}(A,C) \mapsto (A,B,C,D)$

C is the first block entry of $\mathscr{O}_{\mathtt{l}_{\max}+1}(A,C)$ and A is given by

$$\big(\sigma^*\mathscr{O}_{\mathtt{l}_{\max}+1}(A,C)\big)A = \big(\sigma\mathscr{O}_{\mathtt{l}_{\max}+1}(A,C)\big) \quad \text{shift equation}$$

(σ^* removes the last block entry and σ removes the first block entry)

once C and A are known, the system of equations

$$y_d(t) = CA^t x_d(1) + \sum_{\tau=1}^{t-1} CA^{t-1-\tau} B u_d(\tau) + D\delta(t+1), \quad \text{for } t = 1,\dots,\mathtt{l}_{\max}+1$$

is linear in D, B, $x_d(1)$ and can be solved explicitly by using Kronecker products

$w_d \mapsto \mathscr{O}_{\mathtt{l}_{\max}+1}(A,C)$

the columns of $\mathscr{O}_{\mathtt{l}_{\max}+1}(A,C)$ are $\mathtt{n}$ linearly indep. free responses of Σ

under the conditions of FL, such responses can be computed from data

$$\begin{bmatrix} \mathscr{H}_t(u_d) \\ \mathscr{H}_t(y_d) \end{bmatrix} G = \begin{bmatrix} 0 \\ Y_0 \end{bmatrix} \quad \begin{matrix} \leftarrow \text{ zero inputs} \\ \leftarrow \text{ free responses} \end{matrix}$$

in order to obtain lin. indep. free responses, G should be maximal rank

once we have a maximal rank matrix of free responses Y_0

$$Y_0 = \mathscr{O}_{\mathtt{l}_{\max}+1}(A,C) \underbrace{\begin{bmatrix} x_{\text{ini},1} & \cdots & x_{\text{ini},j} \end{bmatrix}}_{X_{\text{ini}}} \quad \text{rank revealing factorization}$$

$\rightsquigarrow \mathscr{O}_{\mathtt{l}_{\max}+1}(A,C)$ and X_{ini}, the factorization fixes the state space basis

$w_d \mapsto \big(x_d(1),\dots,x_d(\mathtt{n}_{\max}+\mathtt{m}+1)\big)$

if the free responses are sequential, *i.e.*, if Y_0 is block-Hankel, then X_{ini} is a state sequence of Σ

computation of sequential free responses is achieved as follows

$$\begin{array}{ccccl} \begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} & G & = & \begin{bmatrix} U_p \\ Y_p \\ 0 \end{bmatrix} & \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{ sequential ini. conditions} \\ \leftarrow \text{ zero inputs} \end{matrix} \\ Y_f & G & = & Y_0 & \end{array} \tag{5}$$

note: now we use the splitting of the data into "past" and "future"

$$Y_0 = \mathscr{O}_{\mathtt{l}_{\max}+1}(A,C) \begin{bmatrix} x_d(1) & \cdots & x_d(\mathtt{n}_{\max}+\mathtt{m}+1) \end{bmatrix} \quad \text{rank revealing factorization}$$

Refinements

- solve (5) efficiently exploiting the Hankel structure
- iteratively compute pieces of Y_0

  $\rightsquigarrow$ iterative algorithm

  - requires smaller persistency of excitation of u_d
  - could be more efficient

    (solves a few smaller systems of eqns instead of a single bigger one)

Relation to other exact identification algorithms

MOESP type algorithms

orthogonal projection of the rows of $\mathscr{H}_{\mathtt{n}_{\max}}(y_d)$ on $\Big(\text{row span}\big(\mathscr{H}_{\mathtt{n}_{\max}}(u_d)\big)\Big)^{\perp}$

$$Y_0 := \mathscr{H}_{\mathtt{n}_{\max}}(y_d)\Pi_{u_d}^{\perp}$$

where

$$\Pi_{u_d}^{\perp} := \Big(I - \mathscr{H}_{\mathtt{n}_{\max}}^{\top}(u_d)\big(\mathscr{H}_{\mathtt{n}_{\max}}(u_d)\mathscr{H}_{\mathtt{n}_{\max}}^{\top}(u_d)\big)^{-1}\mathscr{H}_{\mathtt{n}_{\max}}(u_d)\Big)$$

observe that $\Pi_{u_d}^{\perp}$ is maximal rank and

$$\begin{bmatrix} \mathscr{H}_{\mathtt{n}_{\max}}(u_d) \\ \mathscr{H}_{\mathtt{n}_{\max}}(y_d) \end{bmatrix} \Pi_{u_d}^{\perp} = \begin{bmatrix} 0 \\ Y_0 \end{bmatrix}$$

$\implies$ the orthogonal projection computes free responses

Comments

- $T - \mathtt{n}_{\max} + 1$ free responses are computed via the orth. proj. while $\mathtt{n}_{\max}$ such responses suffice for the purpose of exact identification
- the orth. proj. is a geometric operation, whose system theoretic meaning is not revealed
- the condition for $\operatorname{rank}(Y_0) = \mathtt{n}$, given in the MOESP literature,

  $$\operatorname{rank}\left(\begin{bmatrix} X_{\text{ini}} \\ \mathscr{H}_{\mathtt{n}_{\max}}(u_d) \end{bmatrix}\right) = \mathtt{n} + \mathtt{n}_{\max}\mathtt{m}$$

  is not verifiable from the data (u_d, y_d) $\implies$ can not be checked whether the computation gives $\mathscr{O}(A,C)$, *c.f.*, p.e. condition of FL

N4SID-type algorithms

consider the splitting of the data into "past" and "future"

$$\mathscr{H}_{2\mathtt{n}_{\max}}(u_\mathrm{d}) =: \begin{bmatrix} U_\mathrm{p} \\ U_\mathrm{f} \end{bmatrix}, \qquad \mathscr{H}_{2\mathtt{n}_{\max}}(y_\mathrm{d}) =: \begin{bmatrix} Y_\mathrm{p} \\ Y_\mathrm{f} \end{bmatrix}$$

with $\operatorname{row\,dim}(U_\mathrm{p}) = \operatorname{row\,dim}(U_\mathrm{f}) = \operatorname{row\,dim}(Y_\mathrm{p}) = \operatorname{row\,dim}(Y_\mathrm{f}) = \mathtt{n}_{\max}$ and let

$$W_\mathrm{p} := \begin{bmatrix} U_\mathrm{p} \\ Y_\mathrm{p} \end{bmatrix}$$

the key step of the N4SID algorithms is the oblique projection of the rows of Y_f along $\operatorname{row\,span}(U_\mathrm{f})$ onto $\operatorname{row\,span}(W_\mathrm{p})$

$$Y_0 := Y_\mathrm{f} /_{U_\mathrm{f}} W_\mathrm{p} := Y_\mathrm{f} \underbrace{\begin{bmatrix} W_\mathrm{p}^\top & U_\mathrm{f}^\top \end{bmatrix} \begin{bmatrix} W_\mathrm{p} W_\mathrm{p}^\top & W_\mathrm{p} U_\mathrm{f}^\top \\ U_\mathrm{f} W_\mathrm{p}^\top & U_\mathrm{f} U_\mathrm{f}^\top \end{bmatrix}^+ \begin{bmatrix} W_\mathrm{p} \\ 0 \end{bmatrix}}_{\Pi_\mathsf{obl}}$$

N4SID-type algorithms

observe that

$$\begin{bmatrix} W_\mathrm{p} \\ U_\mathrm{f} \\ Y_\mathrm{f} \end{bmatrix} \Pi_\mathsf{obl} = \begin{bmatrix} W_\mathrm{p} \\ 0 \\ Y_0 \end{bmatrix}$$

(in fact Π_obl is the least-norm, least-squares solution)

$\Longrightarrow$ the oblique projection computes sequential free responses

Comments

- $T - 2\mathtt{n}_{\max} + 1$ sequential free responses are computed via the oblique projection while $\mathtt{n}_{\max} + \mathtt{m} + 1$ such responses suffice for exact ident.
- the oblique proj. is a geometric operation, whose system theoretic meaning is not revealed
- the conditions for $\operatorname{rank}(Y_0) = \mathtt{n}$, given in the N4SID literature,

  1. u_d persistently exciting of order $2\mathtt{n}_{\max}$ and
  2. $\operatorname{row\,span}(X_\mathrm{ini}) \cap \operatorname{row\,span}(U_\mathrm{f}) = \{0\}$

  are not verifiable from the data $(u_\mathrm{d}, y_\mathrm{d})$

Summary for the exact identification algorithms part

- we gave system theoretic interpretation of the orth. and oblique proj.
- MOESP and N4SID alg. are computationally inefficient; more than what is necessary for exact ident. is computed $\rightsquigarrow$ cheaper algorithms
- the FL gives sharp conditions for identifiability, verifiable from the data

  $\rightsquigarrow$ our alg. might be applicable in cases when the classical alg. are not
- we clarified the role of the splitting: the "past" assigns the initial conditions and in the "future" a desired response is computed

  $\Longrightarrow$ the "past" should be chosen at least $\mathtt{l}$ samples long and the length of the "future" is free as long as the p.e. condition is satisfied

Approximate identification

Motivation and approaches

in practice the data is noise corrupted and generated by a complicated system that in general is not in the assumed model class $\mathscr{M} := \mathscr{L}^{\mathtt{w}}_{\mathtt{m},\mathtt{l}}$

$\Longrightarrow$ the data is not exact and the MPUM does not exist in $\mathscr{M}$

$\Longrightarrow$ an approximation is needed in order to find a model in $\mathscr{M}$

approaches for approximate identification:

- modify exact id. alg. by doing approx. lin. algebra (LS, SVD, *etc.*)
- fit the data exactly by a high order model and do model reduction
- minimize an approximation criterion over all $\mathscr{B} \in \mathscr{M}$

Misfit minimization

define the misfit $M(w_\mathrm{d}, \mathscr{B})$ between w_d and $\mathscr{B} \in \mathscr{M}$ as follows

$$M(w_\mathrm{d}, \mathscr{B}) := \min_{\hat{w} \in \mathscr{B}} \|w_\mathrm{d} - \hat{w}\|^2_{\ell_2} \quad \rightsquigarrow \quad \hat{w}^*$$

$\hat{w}^*$ is the projection of w_d on $\mathscr{B}$, *i.e.*, the best ℓ_2 approx. of w_d in $\mathscr{B}$

$\hat{w}^*$ is the smoothed estimate of w_d, given $\mathscr{B}$

our goal is to find the model $\hat{\mathscr{B}}$ that minimizes $M(w_\mathrm{d}, \mathscr{B})$, *i.e.*,

$$\hat{\mathscr{B}} := \arg \min_{\mathscr{B} \in \mathscr{M}} M(w_\mathrm{d}, \mathscr{B}) \qquad \text{global total least squares problem}$$

note: a double minimization problem

Relation to exact identification

modify the given time series as little as possible, so that the MPUM of the modified time series in the model class $\mathscr{M}$ exists

in the errors-in-variables setup, *i.e.*, when the data is generated according to the model

$$w_\mathrm{d} = \bar{w} + \tilde{w}, \qquad \text{where} \quad \bar{w} \in \bar{\mathscr{B}} \in \mathscr{M} \quad \text{and} \quad \tilde{w} \sim \mathrm{N}(0, I)$$

$\hat{\mathscr{B}}$ is the maximum likelihood estimator of the true model $\bar{\mathscr{B}}$

$\hat{\mathscr{B}}$ is a consistent estimator of the true model $\bar{\mathscr{B}}$, *i.e.*, $\hat{\mathscr{B}} \to \bar{\mathscr{B}}$ as $T \to \infty$

Computation of the misfit

misfit computation: given w_{d} and $\mathscr{B} \in \mathscr{M}$, $\min_{\hat{w}\in\mathscr{B}} \|w_{\mathrm{d}} - \hat{w}\|_{\ell_2}^2$

a LS problem; we aim, however, at efficient recursive solution methods

$\mathscr{B} = \mathscr{B}_{\text{iso}}(A,B,C,D)$, the misfit can be computed by dynamic prog.

$\rightsquigarrow$ backwards and forwards processing of the data by time varying systems and solution of Riccati difference equation

$\mathscr{B} = \ker\big(R(\sigma)\big)$ $\rightsquigarrow$ Cholesky factorization of a banded Toeplitz matrix, for which efficient methods based on displacement rank theory exist

Minimization with respect to the model parameters

$\min_{\mathscr{B}\in\mathscr{M}} M(w_{\mathrm{d}},\mathscr{B})$ is a nonlinear least squares problem

$M(w_{\mathrm{d}},\mathscr{B})$, however, is nonconvex with respect to the model parameters

GTLS problem $\equiv$ structured total least squares problem

there are efficient local optim. methods and software for STLS, see

`http://www.esat.kuleuven.ac.be/~imarkovs/stls.html`

this allows to solve routinely GTLS with a few thousands data points

Simulation examples

Examples from DAISY

DAISY — data base for system identification, available from

`http://www.esat.kuleuven.ac.be/~tokka/daisydata.html`

real-life and simulated data for verification and comparison of ident. alg.

- very small ($T = 57$) and very large ($T = 9600$) data sets
- data sets from highly nonlinear phenomena (*e.g.*, wing flutter)
- benchmark problem (*e.g.*, data set # 2)

#	Data set name	parameters T	m	p	l
1	Data of a simulation of the western basin of Lake Erie	57	5	2	1
2	Data of Ethane-ethylene destillation column	90	5	3	1
3	Data from an industrial dryer (Cambridge Control Ltd)	867	3	3	1
4	Wing flutter data	1024	1	1	5
5	Heat flow density through a two layer wall	1680	2	1	2
6	Simulation data of a pH neutralization process	2001	2	1	6
7	Data of a CD-player arm	2048	2	2	1
8	Data from a test setup of an industrial winding process	2500	5	2	2
9	Liquid-saturated steam heat exchanger	4000	1	1	2
10	Data from an industrial evaporator	6305	3	3	1
11	Continuous stirred tank reactor	7500	1	2	1
12	Model of a steam generator at Abbott Power Plant	9600	4	4	1

Simulation setup

the approximations obtained by the following methods are compared:

- `gtls` — the global total least squares method (computed via STLS)
- `pem` — the prediction error method of the Identification Toolbox
- `subid` — robust combined subspace algorithm

numbers in black are for ident. and validation using the whole data set

numbers in blue are for identification on the first 85% of the data and validation on the remaining 15% of the data

$\hat{\mathscr{B}}$ for `detss` and `pem` is the deterministic part of the identified system

ini. approx. for the GTLS optimization method is the estimate of `subid`

#	Data set name	scaled misfit: gtls		pem		subid	
1	Lake Erie	1	1	22.0	9.6	1.5	1.9
2	Destillation	1	1	17.5	14.4	3.1	3.7
3	Industrial dryer	1	1	1.2	1.1	1.2	1.1
4	Wing flutter	1	1.4	2.9	1	1.7	1.5
5	Heat flow	1	1	10.2	10.7	1.9	2.5
6	pH process	1	2.2	2.8	1	1.2	1.4
7	CD-player arm	1	1	1.4	1.4	1.1	1.2
8	Winding process	1	1	2.8	2.6	1.6	1.5
9	Exchanger	1	1	8.1	6.9	1.9	1.6
10	Evaporator	1	1	1.7	1.7	1.6	1.5
11	Tank reactor	1	1	51.5	39.0	2.3	1.6
12	Generator	1	1	3.3	3.1	2.4	2.6

100/100 — identification/validation 85/15 — identification/validation

#	Data set name	scaled exec. time: gtls		pem		subid	
1	Lake Erie	2.3	2.4	6.4	9.6	1	1
2	Destillation	5.7	4.4	19.7	15.8	1	1
3	Industrial dryer	22.5	19.8	20.8	19.7	1	1
4	Wing flutter	2.4	2.3	23.4	12.8	1	1
5	Heat flow	4.5	3.5	36.6	31.4	1	1
6	pH process	5.3	4.7	22.7	36.4	1	1
7	CD-player arm	6.2	13.7	38.2	34.5	1	1
8	Winding process	48.1	41.8	64.0	46.7	1	1
9	Exchanger	5.4	5.1	23.5	37.8	1	1
10	Evaporator	93.0	87.0	133	111	1	1
11	Tank reactor	32.3	29.0	124	118	1	1
12	Generator	288	244	205	207	1	1

100/100 — identification/validation 85/15 — identification/validation

Conclusions

Conclusions

- the ARMAX problem is invariably considered in the sys. id. literature

  we discussed two less usual deterministic identification problems:

  1. exact identification and
  2. approximate identification

- exact identification is interesting and nontrivial; it has two parts:

  1. check if w_{d} completely specifies $\mathscr{B}$ $\rightsquigarrow$ FL
  2. find a desired representation of $\mathscr{B}$ from w_{d}

$$w_{\mathrm{d}} \mapsto R(\xi) \qquad w_{\mathrm{d}} \mapsto H \qquad w_{\mathrm{d}} \mapsto (A,B,C,D) \qquad w_{\mathrm{d}} \mapsto \text{balanced}$$

Conclusions

- approximation is needed in order to treat rough data

  the GTLS problem minimizes the global misfit between w_{d} and $\hat{\mathscr{B}}$

  $\Longrightarrow$ the problem has a meaningful and natural interpretation

  in contrast, the PEM minimizes the one step ahead prediction error, which is a measure for local approximation only

  the GTLS approximation has ML interpretation in the EIV setting

- our current work is focused on the ARMAX identification problem, where the FL again plays a key role

- in this talk, we discussed many topics without delving deeper in the related theoretical and algorithmic issues

Thank you!

References

Deterministic identification

1. J. C. Willems. From time series to linear system—Part I. Finite dimensional linear time invariant systems. *Automatica*, 22(5):561–580, 1986.

2. J. C. Willems. From time series to linear system—Part II. Exact modelling. *Automatica*, 22(6):675–694, 1986.

3. J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 2004.

4. I. Markovsky, J. C. Willems, P. Rapisarda, and B. De Moor. Algorithms for deterministic balanced subspace identification. Technical Report 04–13, Dept. EE, K.U. Leuven, 2004.

Global total least squares

5. B. Roorda and C. Heij. Global total least squares modeling of multivariate time series. *IEEE Trans. on Aut. Control*, 40(1):50–63, 1995.

6. I. Markovsky, J. C. Willems, S. Van Huffel, B. De Moor, and R. Pintelon. Application of structured total least squares for system identification and model reduction. Technical Report 04–51, Dept. EE, K.U. Leuven, 2004.

Numerical methods

7. B. Roorda. Algorithms for global total least squares modelling of finite multivariable time series. *Automatica*, 31(3):391–404, 1995.

8. I. Markovsky, S. Van Huffel, and R. Pintelon. Block-Toeplitz/Hankel structured total least squares. *SIAM J. Matrix Anal.*, 2004.

9. I. Markovsky and S. Van Huffel. High-performance numerical algorithms and software for structured total least squares. Technical Report 04–102, Dept. EE, K.U. Leuven, 2003.

Consistency

10. C. Heij and W. Scherrer. Consistency of system identification by global total least squares. *Automatica*, 35:993–1008, 1999.

11. A. Kukush, I. Markovsky, and S. Van Huffel. Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *J. of Stat. Planning and Inference*, 2004.