Overview of research work

Ivan Markovsky

Katholieke Universiteit Leuven

Background

- BS and MS (TU-Sofia, Bulgaria), PhD (K.U.Leuven)

  in electrical engineering with specialization control systems

- 1.5 years at the University of Notre Dame, supervisor M. Lemmon

  research on hybrid systems

- 4 years at K.U.Leuven, promotors S. Van Huffel and B. De Moor

  research on total least squares and systems identification

- collaborators: J. C. Willems (KUL), A. Kukush (Kiev), P. Rapisarda (Maastricht), R. Pintelon (VUB)

PhD thesis

title: Exact and approximate modeling in the behavioral setting

- generalized (weighted and structured) total least squares problems
- fundamental matrix and ellipsoid estimation
- exact system identification—identifiability conditions and algorithms
- identification of a balanced model
- approximate system identification (global total least squares problem)

properties, computational methods, and software

Outline of this presentation

- overview of prior work
  - exact system identification
  - approximate system identification
- current work and future plans
  - misfit vs latency
  - recursive modeling
  - kernel structure of a block-Hankel matrix
  - software package for structured total least squares

Exact identification

given: a vector time series

$$w = \big(w(1), \dots, w(T)\big)$$

generated by an LTI system $\mathscr{B}$

find: the system $\mathscr{B}$ back from the data w

note: the given data is exact and the identified system fits exactly w
the time horizon T is much larger than the order n of $\mathscr{B}$

Algorithms for exact identification

1. $w \mapsto$ difference equation R

$$R_0 w(t) + R_1 w(t+1) + \cdots + R_l w(t+l) = 0, \quad \text{for } t = 1, \dots, T-l$$

2. $w \mapsto$ impulse response H
3. $w \mapsto$ input/state/output representation (A,B,C,D)

3.a. $w \mapsto R \mapsto (A,B,C,D)$ or $w \mapsto H \mapsto (A,B,C,D)$

3.b. $w \mapsto$ observability matrix $\mapsto (A,B,C,D)$

3.c. $w \mapsto$ state sequence $\mapsto (A,B,C,D)$

Persistency of excitation

a condition for solvability of the exact identification problem

definition: the sequence $u = \big(u(1), \dots, u(T)\big)$ is

persistently exciting of order L

if the Hankel matrix

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

is of full row rank

Fundamental Lemma

Let $\mathscr{B}$ be controllable and let $w := (u,y) \in \mathscr{B}|_{[1,T]}$. Then, if u is persistently exciting of order $L+n$, where n is the order of $\mathscr{B}$,

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

$\Longrightarrow$ with $L = l+1$, where l is the lag of $\mathscr{B}$, the FL gives conditions for identifiability, namely "u persistently exciting of order $l+1+n$"

$\Longrightarrow$ under the conditions of the FL, any L samples long trajectory of $\mathscr{B}$ can be obtained as $\mathscr{H}_L(w)g$, for certain g $\rightsquigarrow$ algorithms

Example $w \mapsto$ impulse response H

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

the problem reduces to the one of finding a particular G

$$\left[\begin{array}{c} \mathscr{H}_{l+t}(u) \\ \hline \mathscr{H}_{l+t}(y) \end{array}\right] G = \left[\begin{array}{c} 0 \\ \left[\begin{smallmatrix} I \\ 0 \end{smallmatrix}\right] \\ \hline 0 \\ H \end{array}\right] \quad \begin{array}{l} \leftarrow \; l \text{ zero samples} \\ \leftarrow \; t \text{ samples long impulse} \\ \hline \leftarrow \; l \text{ zero samples} \\ \leftarrow \; t \text{ samples impulse response} \end{array}$$

block algorithm:

1. solve the system of equations in blue for G

2. substitute G in the equations in red $\rightsquigarrow H$

Simulation example $w \mapsto$ impulse response H

$\mathscr{B}$ is of order $n = 4$, lag $l = 2$, with $m = 2$ inputs, and $p = 2$ outputs

w is a trajectory of $\mathscr{B}$ with length $T = 500$

estimation error $e = ||H - \hat{H}||_F$ and execution time for three methods

method	error, e	time, sec.
block algorithm	10^{-14}	0.293
iterative algorithm	10^{-14}	0.066
`impulse`*	0.059	0.584

* from System Identification Toolbox of MATLAB

Summary

- deterministic subspace algorithms are implementations of the FL

  $w \mapsto$ obsv. matrix $\mapsto (A,B,C,D)$ — MOESP-type algorithms

  $w \mapsto$ state sequence $\mapsto (A,B,C,D)$ — N4SID-type algorithms

- the FL reveals the meaning of the oblique and orthogonal projections

  computation of special responses from data

- the FL gives identifiability conditions that are verifiable from w

Approximate modeling

$\bar{\mathscr{B}}$ — "true" (high order) model

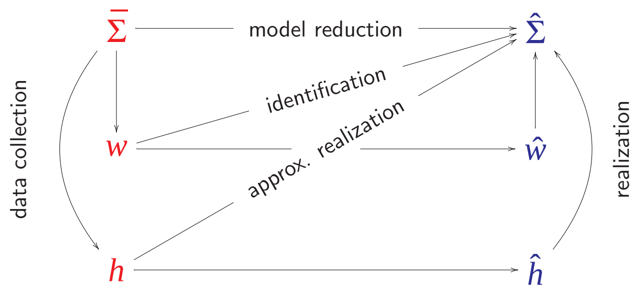
$\hat{\mathscr{B}}$ — approximate (low order) model

w — observed response

h — observed impulse resp.

$\hat{w}$ — response of $\hat{\mathscr{B}}$

$\hat{h}$ — impulse resp. of $\hat{\mathscr{B}}$



Approximate system identification

$\mathscr{M}$ — given model class $\qquad w$ — given time series

the model $\mathscr{B} \in \mathscr{M}$ is viewed as a collection of legitimate time series

the more the model forbids, the less complex and more powerful it is

problem: find a $\hat{\mathscr{B}} \in \mathscr{M}$ that best fits the data according to the

misfit criterion: $\quad M(w,\mathscr{B}) := \min_{\hat{w}\in\mathscr{B}} \|w-\hat{w}\|_{\ell_2}^2 \quad$ (smoothing problem)

the resulting identification problem is:

global total least squares problem: $\quad \min_{\mathscr{B}\in\mathscr{M}} M(w,\mathscr{B})$

A kernel subproblem

- SVD-based methods:

  balanced model reduction, subspace identification, and Kung's alg. use the singular value decomposition in order to find a rank deficient matrix $\mathscr{H}(\hat{w})$ approximating a given full rank matrix $\mathscr{H}(w)$

  note that SVD is suboptimal in terms of the misfit criterion $\|w-\hat{w}\|_{\ell_2}^2$

- structured total least sqaures based method:

  optimal approximation according to the misfit criterion

  need initial approximation (*e.g.*, from SVD-based method)

  iterative improvement of heuristic suboptimal solution

Data sets from DAISY

#	Data set name	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 of a 120 MW power plant	200	5	3	2
4	Heating system	801	1	1	2
5	Data from an industrial dryer (Cambridge Control Ltd)	867	3	3	1
6	Data of a laboratory setup acting like a hair dryer	1000	1	1	5
7	Data of the ball-and-beam setup in SISTA	1000	1	1	2
8	Wing flutter data	1024	1	1	5
9	Data from a flexible robot arm	1024	1	1	4

Data sets from DAISY (cont.)

#	Data set name	T	m	p	l
10	Data of a glass furnace (Philips)	1247	3	6	1
11	Heat flow density through a two layer wall	1680	2	1	2
12	Simulation data of a pH neutralization process	2001	2	1	6
13	Data of a CD-player arm	2048	2	2	1
14	Data from a test setup of an industrial winding process	2500	5	2	2
15	Liquid-saturated steam heat exchanger	4000	1	1	2
16	Data from an industrial evaporator	6305	3	3	1
17	Continuous stirred tank reactor	7500	1	2	1
18	Model of a steam generator at Abbott Power Plant	9600	4	4	1

Setup of the experiment

the approximations obtained by the following methods are compared:

`stls`	—	misfit minimization method
`pem`	—	the prediction error method (Identification Toolbox)
`subid`	—	robust combined subspace algorithm

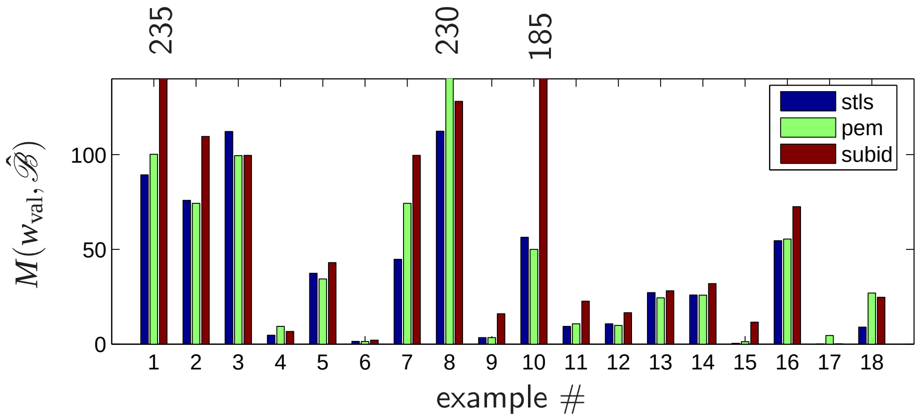
(initial approximation for `stls` and `pem` is the result of `subid`)

a model $\hat{\mathscr{B}}$ is obtained from w_{id} — the first 70% of the data w

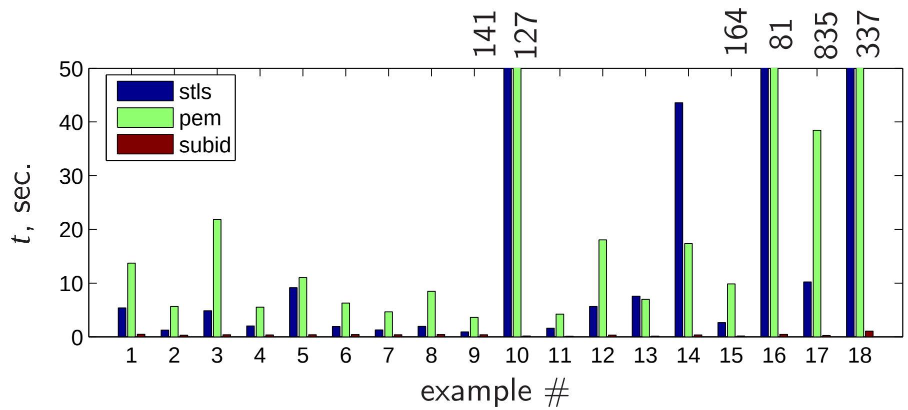
we consider output error identification, *i.e.*, the input is assumed exact

and compare the misfit $M(w_{\text{val}}, \hat{\mathscr{B}})$ on the last 30% of the data w and the execution time for computing $\hat{\mathscr{B}}$

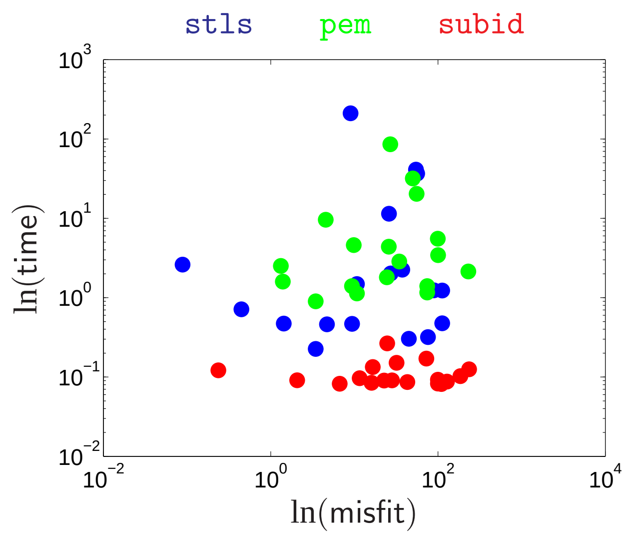
Results — output error



Results — execution time



Results — scatter plot misfit vs time



Summary

- structured total least sqaures is a kernel problem for LTI modeling

  approx. realization, model reduction, system ident., *etc.*

- a single algorithm can solve a large variety of problems

- the software implementation can solve problems with

  a few thousands data points ($T < 10000$), a few outputs ($p < 10$), and a few time lags ($l < 10$)

Misfit vs latency

we measured the data–model discrepancy by the misfit $M(w, \mathscr{B})$

an alternative approach is to introduce an unobserved latent variable e

classically e is assumed to be a random process $\rightsquigarrow$ ARMAX ident.

How do the misfit and latency approaches compare and complement each other?

we aim at methods and algorithms that treat both misfit and latency, thus combining the virtues of both approaches

Recursive modeling

if the model is used while the data is collected, it is necessary to update the model as new data points arrive

moreover, the update should be done under hard time constraints

recursive identification of an LTI system ($\approx$ Berlekamp–Massey algorithm)

A. Antulas and J. C. Willems. A behavioral approach to linear exact modeling. *IEEE-AC*, 1993

M. Kuijper and J. C. Willems. On constructing a shortest linear recurrence relation. *IEEE-AC*, 1997

How to carry out the approximatation recursively?

Kernel structure of a block-Hankel matrix

the classical identification algorithms compute a basis for the left kernel of the Hankel matrix $\mathscr{H}_L(w)$ via a QR factorization followed by SVD

the kernel of $\mathscr{H}_L(w)$ however is highly structured: in the SISO case, it is generated by a single vector $n \in \mathbb{R}^{2(l+1)}$ and its shifts:

$$\operatorname{row\,span}\left(\begin{bmatrix} n_0 & n_1 & \cdots & n_l & 0 & \cdots & 0 \\ 0 & n_0 & n_1 & \cdots & n_l & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & n_0 & n_1 & \cdots & n_l \end{bmatrix}\right) = \operatorname{left\,ker}\left(\mathscr{H}_L(w)\right)$$

we aim to exploit the kernel structure for efficient system ident.

How to carry out the kernel computation recursively?

Software package for structured total least squares

$$\min_X \left(\min_{\hat{w}} \|w - \hat{w}\| \quad \text{subject to} \quad \mathscr{H}_L^\top(\hat{w}) \begin{bmatrix} X \\ -I \end{bmatrix} = 0 \right)$$

the software is written in ANSI C with calls to BLAS and LAPACK

planed extension of the package is to add:

1. regularization
2. diagonal weight matrix in the cost function
3. alternative optimization methods (now Levenberg–Marquardt is used)