Comparison of identification algorithms on the database DAISY

Ivan Markovsky, Jan C. Willems, and Bart De Moor

K.U.Leuven, ESAT-SISTA

Outline

System identification: $w_{\mathrm{d}} \mapsto \widehat{\mathscr{B}} \in \mathscr{M}$

Notation

- $w_{\mathrm{d}} = (u_{\mathrm{d}}, y_{\mathrm{d}})$ — given data, in this talk a vector time series
- $\widehat{\mathscr{B}}$ — to be found model for w_{d}, in this talk an LTI system
- $\mathscr{M}$ — model class, in this talk $\mathscr{L}_{\mathtt{m},\mathtt{n}}$, i.e., LTI systems of bounded complexity: $\leq \mathtt{m}$ inputs and order $\leq \mathtt{n}$

System identification

- defines a mapping $w_{\mathrm{d}} \mapsto \mathscr{B}$
- derives effective algorithms that realize the mapping, and
- develops efficient software that implements the algorithms

Stochastic estimation point of view

Quality certificate: The methods are consistent and efficient under certain specified conditions.

Typical assumptions

- data generated by an ARMAX system
- stationary, white, Gaussian noise

The assumptions imply

- there is a true system $\bar{\mathscr{B}}$ in the model class
- the modeling error $\widetilde{y} := y_{\mathrm{d}} - \widehat{y}$, $\widehat{w} \in \widehat{\mathscr{B}}$ is stationary, etc., stochastic process

however, stochastic identification $\not\equiv$ ARMAX
(errors-in-variables, deterministic approach to latent variables)

Deterministic approximation point of view

- w_d can be generated by a nonlinear time-varying system
- $\widetilde{w} := w_\mathrm{d} - \widehat{w}$, where $\widehat{w} \in \mathscr{B}$, is (most probably) not a stationary stochastic process

The issue is	how to best <u>approximate</u> w_d
rather than	how to best <u>estimate</u> $\mathscr{B} \in \mathscr{M}$.

In the ARMAX setting, the uncertainty is attributed to latent variables that satisfy extra assumptions.

In the approximation setting that we develop, the uncertainly is attributed to the misfit $\widetilde{w} := w_\mathrm{d} - \widehat{w}$.

Resulting identification methods

maximum likelihood estimation in the ARMAX setting
$\rightsquigarrow$ prediction error minimization (PEM):

$$\min_{\widehat{\mathscr{B}}, e} \|e\| \quad \text{subject to} \quad \operatorname{col}(e, u_\mathrm{d}, y_\mathrm{d}) \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{e+m},\mathtt{l}}$$

e — latent variable or prediction error

an alternative approximate identification problem, called global total least squares (GTLS):

$$\min_{\widehat{\mathscr{B}}, \widehat{w}} \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad \widehat{w} \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathtt{m},\mathtt{l}}$$

$w_\mathrm{d} - \widehat{w}$ — data-model misfit

PEM and GTLS minimize (in general) different criteria

Subspace methods

do not minimize an explicit cost function

based on low-rank approximation (SVD) and approximate solution of linear equations (LS)

the main types of subspace methods are:

- N4SID — numerics for system identification
  $w_\mathrm{d} \mapsto x$ (state sequence), then $(w_\mathrm{d}, x) \mapsto (A, B, C, D)$
- MOESP — multivariate output error subspace
  $w_\mathrm{d} \mapsto \mathscr{O}$ (observability matrix), then $(w_\mathrm{d}, \mathscr{O}) \mapsto (A, B, C, D)$
- CVA — canonical variate analysis

Considered identification methods

- `subid`: robust combined subspace algorithm
- `w2x2ss`: N4SID-type algorithm (based on shift-and-cut)
- `cva`: canonical variate analysis method
- `moesp`: multivariable output-error state space
- `pem`: output error identification in the PEM setting
- `gtls`: output error identification using STLS

The first 4 are subspace methods.

The last 2 are optimization based methods.

Database for system identification DAISY

#	Data set name	T	m	p	l
1	Lake Erie	57	5	2	1
2	Distillation column	90	5	3	1
3	Heating system	801	1	1	2
4	Industrial dryer	867	3	3	1
5	Hair dryer	1000	1	1	5
6	Ball-and-beam setup	1000	1	1	2
7	Wing flutter	1024	1	1	5
8	Flexible robot arm	1024	1	1	4
9	Glass furnace	1247	3	6	1
10	Heat flow density	1680	2	1	2
11	pH process	2001	2	1	6
12	CD-player arm	2048	2	2	1
13	Winding process	2500	5	2	2
14	Heat exchanger	4000	1	1	2
15	Industrial evaporator	6305	3	3	1
16	Stirred tank reactor	7500	1	2	1
17	Steam generator	9600	4	4	1

T—# of samples
m—# of inputs
p—# of outputs
l—lag

Simulation setup

w_{d} in all examples is split into identification and validation parts.

"70i/30v" is a short notation for "first 70% of the data is used for identification and the remaining 30% for validation"

A model $\widehat{\mathscr{B}}$ is identified from w_{idt} and is validated on w_{val} by the validation criterion defined next.

Validation criterion: "simulation fit"

Let $\bar{y}$ be the mean of y

$$\bar{y} := \frac{1}{T}\sum_{t=1}^{T} y(t)$$

and define

$$\widehat{y}\big((u,y),\mathscr{B}\big) := \min_{\widehat{y}} \|y-\widehat{y}\| \quad \text{subject to} \quad \operatorname{col}(u,\widehat{y}) \in \mathscr{B}.$$

The simulation fit of w by $\mathscr{B}$ is

$$F(w,\mathscr{B}) := 100 \max\big(0, 1 - \|y-\widehat{y}(w,\mathscr{B})\| / \|y-\bar{y}\|\big).$$

Average fit in % on all datasets

Experiment		subid	w2x2ss	moesp	cva	pem	gtls
	idt	51.18	46.39	55.52	49.79	57.43	68.46
70i/30v	val	32.14	32.34	38.97	33.38	37.77	48.40
	idt	46.34	48.83	53.86	50.78	59.13	68.87
30v/70i	val	36.96	38.15	40.43	37.10	45.17	53.72
	idt	49.14	45.56	55.13	50.88	56.84	68.36
80i/20v	val	30.01	29.75	33.01	31.75	36.17	44.14
	idt	49.47	48.07	54.48	51.90	58.93	68.48
20v/80i	val	46.09	40.81	39.79	39.81	45.28	56.88
	idt	50.92	47.61	54.79	51.25	58.39	68.95
90i/10v	val	40.47	31.46	37.06	35.07	39.48	48.55
	idt	48.16	48.46	53.93	50.71	58.78	69.06
10v/90i	val	45.58	45.13	44.12	39.71	43.62	56.28
Exec. time		0.11	0.05	4.45	5.03	14.79	25.14

The best fits and smallest execution times obtained by subspace and optimization methods are in red.

Summary of the results

- `gtls` achieves the best fit
- among the subspace methods, `moesp` achieves the best fit
- fastest (and perhaps most efficient) is `w2x2ss`
- `pem` achieves worse performance than `gtls` (mainly) due to the imposed stability constraint
- a good fit on w_{idt} does not guarantee a good fit on w_{val}

About the data preprocessing

Data preprocessing

- detrending,
- scaling,
- filtering, etc.

is considered obligatory.

However, an optimal identification method $w_{\text{d}} \mapsto \mathscr{B}$ should necessarily include all preprocessing steps.

The preprocessing operations give extra degrees of freedom (apart from the identification criterion) to influence the result.

Important in practice but rather add hock: requires experience, rule-of-thumb rules, trail and error

An example where preprocessing can be avoided

The simplest type of detrending is removing the mean.

very common, known to improve the results of PEM

However, the trend might be an important aspects of the data generating mechanism.

a model for the trend is an unstable system $\implies$ full model is unstable

If the identified model is not constrained to be stable, the trend is modeled automatically by the identification method.

Thank you