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

Certificates for the methods

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 $\tilde{y} := y_{\mathrm{d}} - \hat{y}$, $\hat{w} \in \widehat{\mathscr{B}}$ is stationary, *etc*, stochastic process.

Approximation point of view

- w_{d} can be anything, *e.g.*, 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 approximate w_{d}
instead of how to best estimate $\mathscr{B} \in \mathscr{M}$.

Considered identification methods

- `subid`: robust combined subspace algorithm
- `w2x2ss`: N4SID-type algorithm
- `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

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

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

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

Average fit in % on all datasets

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

Conclusions

- `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 $w_{\rm idt}$ does not guarantee a good fit on $w_{\rm val}$

Thank you