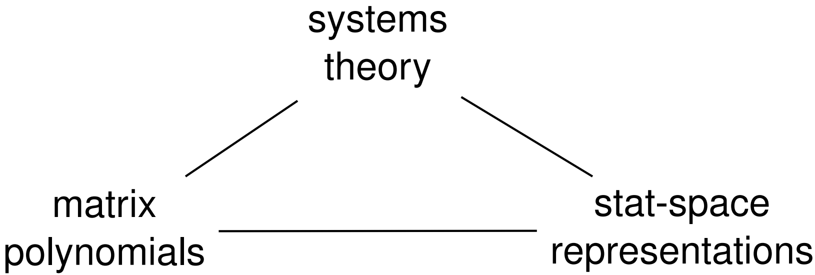
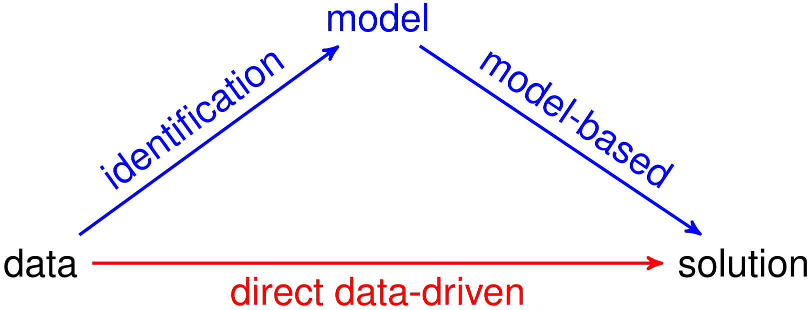
Computations for systems and control without model parameters

Ivan Markovsky

The classical approach is model-based *i.e.*, it is based on model parameters



There are new opportunities and challenges in numerical methods for systems & control



Systems theory, signal processing, and control are going through third paradigm shift

period	paradigm	types of systems
1940–60	classical	SISO transfer funct.
1960–80	modern	MIMO state space
1980–00	behavioral	system as a set
2000–	data-driven	using directly data

New paradigm brings new notion of system and new techniques for solving problems

system	techniques
transfer funct.	Laplace/Z, Fourier transforms
state-space	Lyapunov, Riccati eqn., LMIs
kernel repr.	polynomial algebra
data-driven	numerical linear algebra for structured matrices

Outline

Behavioral approach

Interpolation/approximation of trajectories

Special case: input estimation

Outline

Behavioral approach

Interpolation/approximation of trajectories

Special case: input estimation

We view systems as sets of signals

$w \in (\mathbb{R}^q)^{\mathbb{N}}$ — q-variate discrete-time signal

$\mathscr{B} \subset (\mathbb{R}^q)^{\mathbb{N}}$ — q-variate dynamical model

- linear — $\mathscr{B}$ is a linear subspace of $(\mathbb{R}^q)^{\mathbb{N}}$
- time-invariant — invariant under shifts: $(\sigma w)(t) := w(t+1)$

$w \in \mathscr{B}$ means "w is a trajectory of $\mathscr{B}$"

In practice, we deal with finite signals

restriction of w / $\mathscr{B}$ to finite horizon $[1, T]$

$$w|_T := \big(w(1), \dots, w(T)\big), \quad \mathscr{B}|_T := \{\, w|_T \mid w \in \mathscr{B} \,\}$$

for $w_\text{d} = \big(w_\text{d}(1), \dots, w_\text{d}(T_\text{d})\big)$ and $1 \leq T \leq T_\text{d}$

$$\mathscr{H}_T(w_\text{d}) := \begin{bmatrix} (\sigma^0 w_\text{d})|_T & (\sigma^1 w_\text{d})|_T & \cdots & (\sigma^{T_\text{d}-T} w_\text{d})|_T \end{bmatrix}$$

$w_\text{d} \in \mathscr{B}|_{T_\text{d}}$ — "exact data"

The set of linear time-invariant systems $\mathscr{L}$ has structure characterized by integers

m — number of inputs

n — order ($=$ minimal state dimension)

ℓ — lag ($=$ observability index)

$\mathscr{L}_{(m,\ell,n)}$ — bounded complexity LTI systems

Nonparametric representation of LTI system's finite-horizon behavior

assumptions:

- $w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}}$ — exact offline data
- $\mathscr{B} \in \mathscr{L}_{(m,\ell,n)}$ — bounded complexity LTI system
- informative data, for $T \geq \boldsymbol{\ell}(\mathscr{B})$

$$\operatorname{rank} \mathscr{H}_T(w_{\mathrm{d}}) = mT + n \qquad \text{(GPE)}$$

then, the data-driven representation holds

$$\operatorname{image} \mathscr{H}_T(w_{\mathrm{d}}) = \mathscr{B}|_T \qquad \text{(DDR)}$$

Outline

Behavioral approach

Interpolation/approximation of trajectories

Special case: input estimation

Generic problem: trajectory interpolation and approximation

given:

"data trajectory"	$w_\mathrm{d} \in \mathscr{B}\vert_{T_\mathrm{d}}$
and elements	$w\vert_{I_\text{given}}$
of a trajectory	$w \in \mathscr{B}\vert_T$

($w\vert_{I_\text{given}}$ selects the elements of w, specified by I_given)

aim:

$$\text{minimize} \quad \text{over } \widehat{w} \quad \| w\vert_{I_\text{given}} - \widehat{w}\vert_{I_\text{given}} \|$$
$$\text{subject to} \quad \widehat{w} \in \mathscr{B}\vert_T$$

$$\widehat{w} = \mathscr{H}_T(w_\mathrm{d}) \big(\mathscr{H}_T(w_\mathrm{d})\vert_{I_\text{given}} \big)^+ w\vert_{I_\text{given}} \qquad \text{(SOL)}$$

I. Markovsky and F. Dörfler. "Data-driven dynamic interpolation and approximation". In: *Automatica* 135 (2022), p. 110008

"In linear systems theory, results are either trivial or wrong." *P. Antsaklis*

"A good essay has to be surprising." *P. Graham*

Special cases

simulation

- given data: initial condition and input
- to-be-found: output (exact interpolation)

smoothing

- given data: noisy trajectory
- to-be-found: ℓ_2-optimal approximation

tracking control

- given data: to-be-tracked trajectory
- to-be-found: ℓ_2-optimal approximation

Generalizations

multiple data trajectories $w_d^1, \dots, w_d^N$

$$\widehat{\mathscr{B}}|_L = \text{image}\ \underbrace{\begin{bmatrix} \mathscr{H}_L(w_d^1) & \cdots & \mathscr{H}_L(w_d^N) \end{bmatrix}}_{\text{mosaic-Hankel matrix}}$$

w_d not exact / noisy

maximum-likelihood estimation
$\rightsquigarrow$ Hankel structured low-rank approximation/completion
nuclear norm and ℓ_1-norm relaxations
$\rightsquigarrow$ nonparametric, convex optimization problems

nonlinear systems

results for special classes of nonlinear systems:
Volterra, Wiener-Hammerstein, bilinear, LPV, ...

Outline

Behavioral approach

Interpolation/approximation of trajectories

Special case: input estimation

Input estimation is an old problem, however new results are still being published

S. Gillijns and B. De Moor. "Unbiased minimum-variance input and state estimation for linear discrete-time systems". In: *Automatica* 43.1 (2007), pp. 111–116

M. Abooshahab et al. "Simultaneous input & state estimation, singular filtering and stability". In: *Automatica* 137 (2022), p. 110017

G. Gakis and M. Smith. "Simultaneous input and state estimation for systems with arbitrary inherent delay". In: *IEEE Conference on Decision and Control*. 2024, pp. 2715–2720

Problem statement and data-driven solution

given, $w_{\mathrm{d,ext}} := \begin{bmatrix} e_{\mathrm{d}} \\ w_{\mathrm{d}} \end{bmatrix} \in \mathscr{B}|_{T_{\mathrm{d}}}$ and $w \in \Pi_w \mathscr{B}|_T$

find e, such that $\begin{bmatrix} e \\ w \end{bmatrix} \in \mathscr{B}|_T$

solution: $\widehat{e} = \big(\Pi_e \mathscr{H}_T(w_{\mathrm{d,ext}})\big)\big(\Pi_w \mathscr{H}_T(w_{\mathrm{d,ext}})\big)^+ w$

fact: exact recovery $\widehat{e} = e$, assuming

$$\operatorname{rank} \Pi_w \mathscr{H}_T(w_{\mathrm{d,ext}}) = \operatorname{rank} \mathscr{H}_T(w_{\mathrm{d,ext}})$$

Summary

assuming $\quad \operatorname{rank} \mathscr{H}_L(w_\mathrm{d}) = \mathbf{m}(\mathscr{B})L + \mathbf{n}(\mathscr{B})$

$\mathscr{B}|_L = \operatorname{image} \mathscr{H}_L(w_\mathrm{d})$ holds and

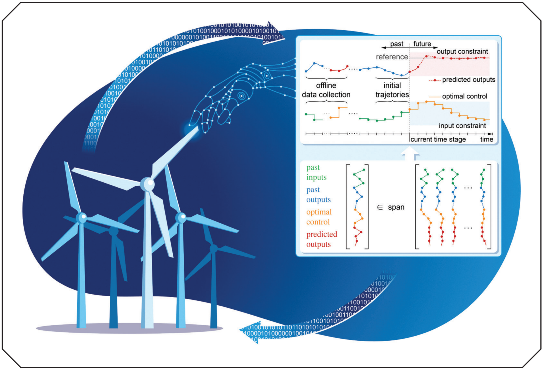
replaces parametric representations

data-driven solution $=$ model-based solution

methods exploiting the structure are needed

Data-Driven Control Based on the Behavioral Approach

FROM THEORY TO APPLICATIONS IN POWER SYSTEMS



SONIA MONTI

IVAN MARKOVSKY, LINBIN HUANG, and FLORIAN DÖRFLER

Outline

Dealing with noise

Empirical validation

The data w_d being exact vs inexact / "noisy"

w_d exact and informative

- "systems theory" problems
- image $\mathscr{H}_L(w_\mathrm{d})$ is nonparametric finite-horizon model
- data-driven solution $=$ model-based solution

w_d inexact, due to noise and/or nonlinearities

- naive approach: apply the solution (SOL) for exact data
- rigorous: assume noise model $\rightsquigarrow$ ML estimation problem
- heuristics: convex relaxations of the ML estimator

The maximum-likelihood estimation problem in the errors-in-variables setup is nonconvex

errors-in-variables setup: $w_{\rm d} = \overline{w}_{\rm d} + \widetilde{w}_{\rm d}$

- $\overline{w}_{\rm d}$ — true data, $\overline{w}_{\rm d} \in \mathscr{B}|_{T_{\rm d}}$, $\mathscr{B} \in \mathscr{L}_c^q$
- $\widetilde{w}_{\rm d}$ — zero mean, white, Gaussian measurement noise

ML problem: given $w_{\rm d}$, c, and $w|_{I_{\rm given}}$

$$\begin{aligned} \text{minimize}_{g} \quad & \|w|_{I_{\rm given}} - \mathscr{H}_T(\widehat{w}_{\rm d}^*)|_{I_{\rm given}} g\| \\ \text{subject to} \quad & \widehat{w}_{\rm d}^* = \arg\min_{\widehat{w}_{\rm d}, \widehat{\mathscr{B}}} \quad \|w_{\rm d} - \widehat{w}_{\rm d}\| \\ & \qquad \text{subject to} \quad \widehat{w}_{\rm d} \in \widehat{\mathscr{B}}|_{T_{\rm d}} \text{ and } \widehat{\mathscr{B}} \in \mathscr{L}_c^q \end{aligned}$$

The ML estimation problem is equivalent to Hankel structured low-rank approximation

$$\begin{aligned}
&\underset{g}{\text{minimize}} \quad \| w|_{I_{\text{given}}} - \mathscr{H}_T(\widehat{w}_{\text{d}}^*)|_{I_{\text{given}}} g \| \\
&\text{subject to} \quad \widehat{w}_{\text{d}}^* = \arg\min_{\widehat{w}_{\text{d}}, \widehat{\mathscr{B}}} \quad \| w_{\text{d}} - \widehat{w}_{\text{d}} \| \\
&\qquad\qquad\qquad \text{subject to} \quad \widehat{w}_{\text{d}} \in \widehat{\mathscr{B}}|_{T_{\text{d}}} \text{ and } \widehat{\mathscr{B}} \in \mathscr{L}_c^q
\end{aligned}$$

$$\Updownarrow$$

$$\begin{aligned}
&\underset{g}{\text{minimize}} \quad \| w|_{I_{\text{given}}} - \mathscr{H}_T(\widehat{w}_{\text{d}}^*)|_{I_{\text{given}}} g \| \\
&\text{subject to} \quad \widehat{w}_{\text{d}}^* = \arg\min_{\widehat{w}_{\text{d}}} \quad \| w_{\text{d}} - \widehat{w}_{\text{d}} \| \\
&\qquad\qquad\qquad \text{subject to} \quad \text{rank}\, \mathscr{H}_{\ell+1}(\widehat{w}_{\text{d}}) \leq (\ell+1)m + n
\end{aligned}$$

Solution methods

local optimization (on a manifold)

- choose a parametric representation of $\widehat{\mathscr{B}}(\theta)$
- optimize over $\widehat{w}$, $\widehat{w_{\text{d}}}$, and θ
- depends on the initial guess

convex relaxation based on the nuclear norm

$$\text{minimize} \quad \text{over } \widehat{w}_{\text{d}} \text{ and } \widehat{w} \quad \| w|_{I_{\text{given}}} - \widehat{w}|_{I_{\text{given}}} \| + \| w_{\text{d}} - \widehat{w}_{\text{d}} \| + \gamma \cdot \left\| \begin{bmatrix} \mathscr{H}_{\Delta}(\widehat{w}_{\text{d}}) & \mathscr{H}_{\Delta}(\widehat{w}) \end{bmatrix} \right\|_*$$

convex relaxation based on ℓ_1-norm (LASSO)

$$\text{minimize} \quad \text{over } g \quad \| w|_{I_{\text{given}}} - \mathscr{H}_T(w_{\text{d}})|_{I_{\text{given}}} g \| + \lambda \| g \|_1$$

Outline

Dealing with noise

Empirical validation

Empirical validation on real-life datasets

	data set name	T_d	m	p
1	Air passengers data	144	0	1
2	Distillation column	90	5	3
3	pH process	2001	2	1
4	Hair dryer	1000	1	1
5	Heat flow density	1680	2	1
6	Heating system	801	1	1

G. Box, and G. Jenkins. Time Series Analysis: Forecasting and Control, Holden-Day, 1976

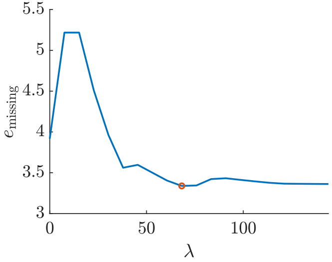
B. De Moor, et al.*DAISY: A database for identification of systems. Journal A, 38:4–5, 1997*

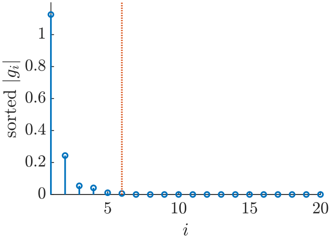
ℓ_1-norm regularization with optimized λ achieves the best performance

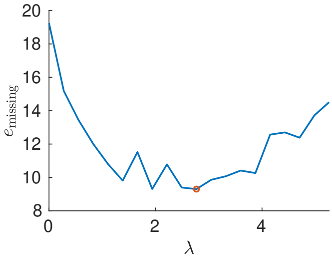
$$e_{\text{missing}} := \frac{\left\| w|_{I_{\text{missing}}} - \widehat{w}|_{I_{\text{missing}}} \right\|}{\left\| w|_{I_{\text{missing}}} \right\|} \, 100\%$$

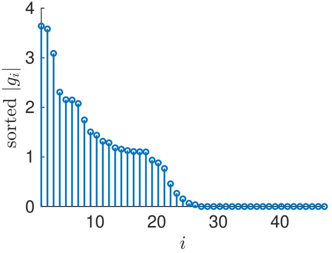
	data set name	naive	ML	LASSO
1	Air passengers data	3.9	fail	3.3
2	Distillation column	19.24	17.44	9.30
3	pH process	38.38	85.71	12.19
4	Hair dryer	12.35	8.96	7.06
5	Heat flow density	7.16	44.10	3.98
6	Heating system	0.92	1.35	0.36

Tuning of λ and sparsity of g (datasets 1, 2)

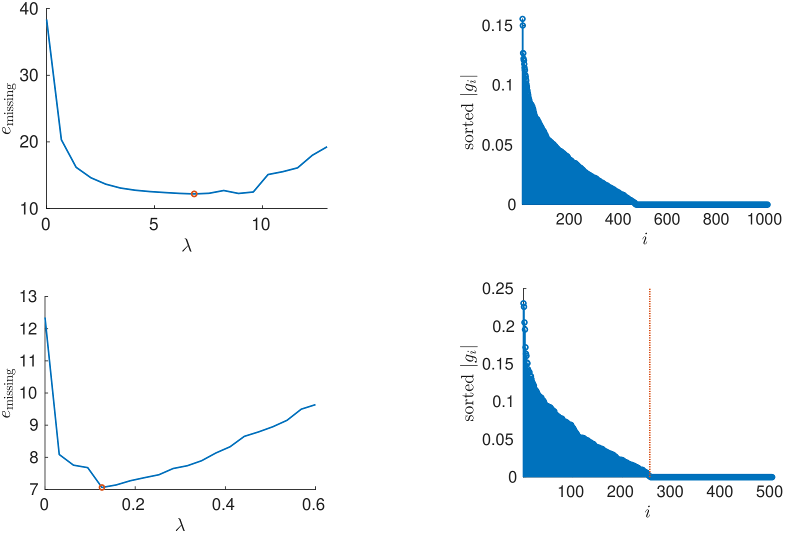








Tuning of λ and sparsity of g (datasets 3, 4)



Tuning of λ and sparsity of g (datasets 5, 6)