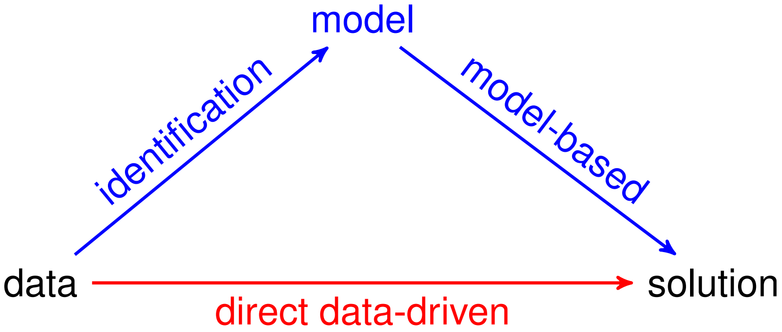
Computations for systems and control without model parameters

Ivan Markovsky

ICREA

CIMNE
INTERNATIONAL CENTRE
FOR NUMERICAL METHODS
IN ENGINEERING

The current trend in systems & control is bypassing parametric model identification



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

However, existing computational methods are still based on model parameters

systems theory

polynomials

state-space

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

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

The Behavioral Toolbox implements direct data-driven solution method

embraces the behavioral approach

deals seamlessly with general LTI systems

does computations without (A, B, C, D) and $H(z)$

Outline

Data-driven representation of LTI systems

Interpolation and approximation of trajectories

Nonparametric frequency response estimation

Direct data-driven fault detection

Outline

Data-driven representation of LTI systems

Interpolation and approximation of trajectories

Nonparametric frequency response estimation

Direct data-driven fault detection

A dynamical system $\mathscr{B}$ is a set of signals

$$w \in \mathscr{B} \quad \leftrightarrow \quad \text{"}w \text{ is trajectory of } \mathscr{B}\text{"}$$
$$\leftrightarrow \quad \text{"}\mathscr{B} \text{ is exact model for } w\text{"}$$

$\mathscr{B}$ is linear system $:\Longleftrightarrow$ $\mathscr{B}$ is subspace

$\mathscr{B}$ is time-invariant $:\Longleftrightarrow$ $\sigma\mathscr{B} = \mathscr{B}$

$(\sigma w)(t) := w(t+1)$ — shift operator

$\sigma\mathscr{B} := \{\sigma w \mid w \in \mathscr{B}\}$

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

the dimension of $\mathscr{B} \in \mathscr{L}$ is determined by

$\mathbf{m}(\mathscr{B})$ — number of inputs

$\mathbf{n}(\mathscr{B})$ — order ($=$ minimal state dimension)

$\boldsymbol{\ell}(\mathscr{B})$ — lag ($=$ observability index)

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

Identifiability: $w_\mathrm{d} \in \mathscr{B}$ specifies $\mathscr{B} \in \mathscr{L}$

define $\quad \widehat{\mathscr{B}} := \operatorname{span}\{\, w_\mathrm{d}, \sigma w_\mathrm{d}, \sigma^2 w_\mathrm{d}, \dots \}$

$\widehat{\mathscr{B}} \in \mathscr{L}$ and $\widehat{\mathscr{B}} \subseteq \mathscr{B}$

identifiability condition: $\quad \widehat{\mathscr{B}} = \mathscr{B}$

J.C. Willems, From time series to linear systems.
Part II, Exact modelling, Automatica, 22(675–694), 1986

We aim to obtain finite horizon results

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

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

$$\dim \mathscr{B}|_T = \mathbf{m}(\mathscr{B})\, T + \mathbf{n}(\mathscr{B}), \quad \text{for all } T \geq \boldsymbol{\ell}(\mathscr{B})$$

for $\mathscr{B}, \mathscr{B}' \in \mathscr{L}$, $\quad \mathscr{B} = \mathscr{B}' \quad$ if and only if

$$\mathscr{B}|_T = \mathscr{B}'|_T, \text{ for } T = \max\{\, \boldsymbol{\ell}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}') \,\} + 1$$

Shifting and cutting w_d leads to Hankel matrix

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

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

define $\quad \widehat{\mathscr{B}}_T := \operatorname{image} \mathscr{H}_T(w_\mathrm{d})$

then, $\quad \widehat{\mathscr{B}}_T \subseteq \mathscr{B}|_T$

Identifiability condition that is verifiable from $w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}}$ and $\big(\mathbf{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \mathbf{n}(\mathscr{B})\big)$

$$\begin{aligned}
\widehat{\mathscr{B}} = \mathscr{B} \quad &\iff \quad \widehat{\mathscr{B}}|_{\boldsymbol{\ell}(\mathscr{B})+1} = \mathscr{B}|_{\boldsymbol{\ell}(\mathscr{B})+1} \\
&\iff \quad \dim \widehat{\mathscr{B}}|_{\boldsymbol{\ell}(\mathscr{B})+1} = \dim \mathscr{B}|_{\boldsymbol{\ell}(\mathscr{B})+1}
\end{aligned}$$

$\mathscr{B}$ is identifiable from $w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}}$ if and only if

$$\operatorname{rank} \mathscr{H}_{\boldsymbol{\ell}(\mathscr{B})+1}(w_{\mathrm{d}}) = \big(\boldsymbol{\ell}(\mathscr{B}) + 1\big)\mathbf{m}(\mathscr{B}) + \mathbf{n}(\mathscr{B})$$

Nonparametric repr. $\mathscr{B}|_T = \text{image}\, \mathscr{H}_T(w_\text{d})$

$\widehat{\mathscr{B}}_T \subseteq \mathscr{B}|_T$, $T \geq \boldsymbol{\ell}(\mathscr{B})$, equality holds if and only if

$$\text{rank}\, \mathscr{H}_T(w_\text{d}) = T\mathbf{m}(\mathscr{B}) + \mathbf{n}(\mathscr{B}) \qquad \text{(GPE)}$$

sufficient conditions ("fundamental lemma"):

1. $w_\text{d} = \begin{bmatrix} u_\text{d} \\ y_\text{d} \end{bmatrix}$
2. $\mathscr{B}$ controllable
3. $\mathscr{H}_{T+\mathbf{n}(\mathscr{B})}(u_\text{d})$ full row rank

J.C. Willems et al., A note on persistency of excitation
Systems & Control Letters, (54)325–329, 2005

Outline

Data-driven representation of LTI systems

Interpolation and approximation of trajectories

Nonparametric frequency response estimation

Direct data-driven fault detection

Data-driven interpolation is missing data recovery

given:

$$w_{\mathrm{d}} \in \mathscr{B}|_{T_{\mathrm{d}}} \quad — \quad \text{“data” trajectory}$$

$$w|_{I_{\mathrm{given}}} \quad — \quad \text{partially specified trajectory}$$

($w|_{I_{\mathrm{given}}}$ selects the elements of w, specified by I_{given})

find:

$$\widehat{w} \in \mathscr{B}|_{T}, \quad \text{such that} \quad \widehat{w}|_{I_{\mathrm{given}}} = w|_{I_{\mathrm{given}}}$$

Solution may not exist

A1: $\operatorname{rank} \mathscr{H}_T(w_\text{d}) = \mathbf{m}(\mathscr{B})T + \mathbf{n}(\mathscr{B})$

A2: $w|_{I_\text{given}}$ has exact completion $w \in \mathscr{B}|_T$

$$\operatorname{rank} \begin{bmatrix} \mathscr{H}_T(w_\text{d})|_{I_\text{given}} & w|_{I_\text{given}} \end{bmatrix} = \operatorname{rank} \mathscr{H}_T(w_\text{d})|_{I_\text{given}}$$

($M|_{I_\text{given}}$ selects the submatrix of M with rows in I_given)

A1 + A2 $\implies$ exact solution exists

Solution may not be unique

when recovered solution is exact, *i.e.*, $\widehat{w} = w$?

A3: there are "enough" given samples I_{given}

$$\operatorname{rank} \mathscr{H}_T(w_\text{d})|_{I_{\text{given}}} = \operatorname{rank} \mathscr{H}_T(w_\text{d})$$

A2 and A3 are verifiable from the data
A1 requires in addition $\mathbf{m}(\mathscr{B}), \boldsymbol{\ell}(\mathscr{B}), \mathbf{n}(\mathscr{B})$

Data-driven interpolation method: solve system of linear equations

there is g, such that $w = \mathscr{H}_T(w_\mathrm{d})g$

method:

1. solve $\quad w|_{I_\text{given}} = \mathscr{H}_T(w_\mathrm{d})|_{I_\text{given}} g$
2. define $\quad \widehat{w} := \mathscr{H}_T(w_\mathrm{d})g$

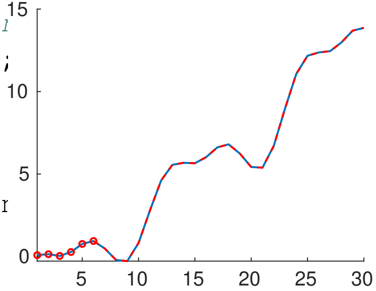
Simulation is special case of interpolation

```
%% interpolation points
t_given = 1:n;
w_given = rand(n, 1);

%% data-driven interpolatio
g = H(t_given, :) \ w_given;
w = H * g;

%% check the result
ws = initial(B, w_given, T-r
norm(ws - w(n:T)) % = 0?

%% results
plot(w), hold on,
plot(n:T, ws, 'r--')
plot(t_given, w_given, 'rO')
```



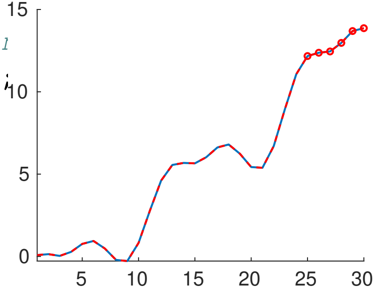
Simulation with “terminal conditions”

```
%% interpolation points
t_given = T-n+1:T;
w_given = w(t_given);

%% data-driven interpolation
g = H(t_given, :) \ w_given;
w_final = H * g;

%% check the result
norm(w - w_final) % = 0?

%% results
plot(w_final), hold on,
plot(w, 'r--')
plot(t_given, w_given, 'rO')
```

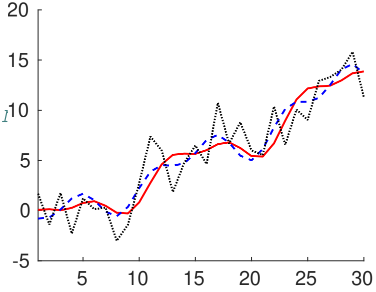


Data-driven approximation (errors-in-variables Kalman smoothing)

```
%% noisy trajectory
w0 = w; % exact trajectory
w  = w0 + 2 * randn(T, 1);

%% data-driven approximation
wh = H * pinv(H) * w;

%% results
plot(w0, 'r-'), hold on
plot(wh, 'b--')
plot(w,  ':k')
```



Generalizations

w_d not exact / noisy

maximum-likelihood estimation
$\rightsquigarrow$ Hankel structured low-rank approximation/completion
parametric, non-convex optimization problem

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, ...

efficient / real-time computation

non-parametric version of the Kalman filter

Outline

Data-driven representation of LTI systems

Interpolation and approximation of trajectories

Nonparametric frequency response estimation

Direct data-driven fault detection

Problem formulation

given: "data" trajectory $(u_\mathrm{d}, y_\mathrm{d}) \in \mathscr{B}|_{T_\mathrm{d}}$ and $z \in \mathbb{C}$

find: $H(z)$, where H is the transfer function of $\mathscr{B}$

Direct data-driven solution

we are interested in trajectory

$$w = \begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} \exp_z \\ \widehat{H} \exp_z \end{bmatrix} \in \mathscr{B}, \quad \text{where } \exp_z(t) := z^t$$

using the data-driven representation, we have

$$\begin{bmatrix} \mathscr{H}_T(u_\mathrm{d}) \\ \mathscr{H}_T(y_\mathrm{d}) \end{bmatrix} g = \begin{bmatrix} \mathbf{z} \\ \widehat{H}\mathbf{z} \end{bmatrix}, \quad \text{where } \mathbf{z} := \begin{bmatrix} z^1 \\ \vdots \\ z^T \end{bmatrix}$$

which leads to the system

$$\begin{bmatrix} 0 & \mathscr{H}_T(u_\mathrm{d}) \\ -\mathbf{z} & \mathscr{H}_T(y_\mathrm{d}) \end{bmatrix} \begin{bmatrix} \widehat{H} \\ g \end{bmatrix} = \begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix} \tag{SYS}$$

Solution method: solve (SYS) for $\widehat{H}$

under (GPE) with $T \geq \ell + 1$, $\widehat{H} = H(z)$

trivial generalization to

- multivariable systems
- multiple data trajectories $\{ w_\text{d}^1, \dots, w_\text{d}^N \}$
- evaluation of $H(z)$ at multiple points in $\{ z_1, \dots, z_K \} \in \mathbb{C}^K$

Comparison with classical nonparametric frequency response estimation methods

ignored initial/terminal conditions $\rightsquigarrow$ *leakage*

DFT grid $\rightsquigarrow$ limited *frequency resolution*

improvements by windowing and interpolation

- the leakage is not eliminated
- the methods involve *hyper-parameters*

Generalization of (SYS) to noisy data

preprocessing: rank-$mT+n$ approx. of $\mathscr{H}_T(w_\mathrm{d})$

- hyper-parameters $T \geq \ell+1$ and n
- if the approximation preserves the Hankel structure, the method is maximum-likelihood in the EIV setting

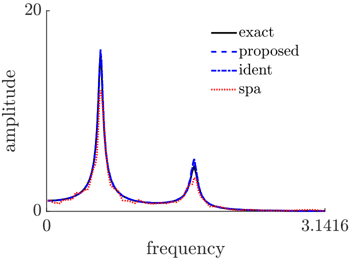
regularization with $\|g\|_1$

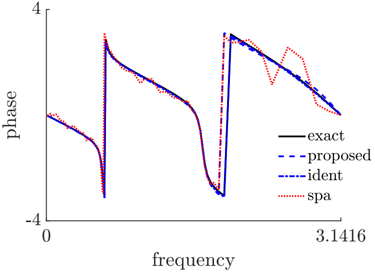
- hyper-parameter: the 1-norm regularization parameter

Example: EIV setup with 4th order system

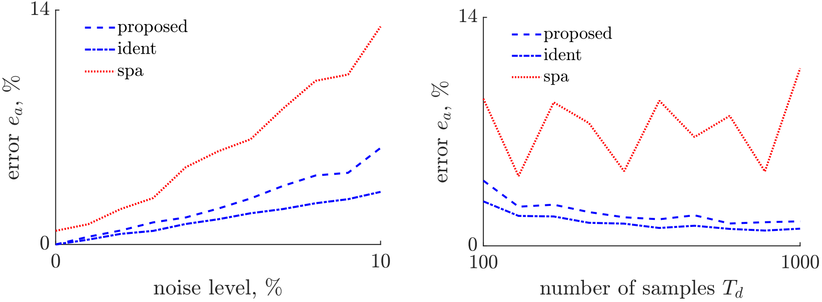
the proposed method is compared with

- `ident` — parametric maximum-likelihood estimator
- `spa` — nonparameteric estimator with Welch filter





Monte-Carlo simulation over different noise levels and number of samples



$$e_a := 100\% \cdot |(|\overline{H}_z| - |\widehat{H}_z|)| \;/\; |\overline{H}_z|$$

Outline

The problem considered is to detect abnormal operation based on observed data

prior information about data-generating system

model-based vs direct data-driven methods

observed data collected offline and online

- dedicated experiment — known excitation signal
- “normal” operation — unknown excitation signal

We consider three data collection scenarios

free response / transient data

forced response with known excitation

forced response with unknown excitation

Recall the nonparametric representation of an LTI system's finite-horizon behavior

assumptions:

- $w_\text{d} \in \mathscr{B}|_{T_\text{d}}$ — exact offline data
- $\mathscr{B} \in \mathscr{L}_{(m,\ell,n)}$ — bounded complexity LTI system
- for $T \geq \boldsymbol{\ell}(\mathscr{B})$, $\operatorname{rank} \mathscr{H}_T(w_\text{d}) = mT + n$ — informative data

then, the data-driven representation holds

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

The fault detection criterion is the distance from online data w to system's behavior $\mathscr{B}$

$\text{dist}(w, \mathscr{B}) := \min_{\widehat{w} \in \mathscr{B}|_T} \|w - \widehat{w}\|$

under the assumptions, using (DDR), we have

$$\text{dist}(w, \mathscr{B}) = \|w - \mathscr{H}_T(w_\text{d}) \mathscr{H}_T^+(w_\text{d}) w\|$$

direct data-driven computation of the distance

The fault detection method has offline and online steps

offline: using w_{d}, find orthonormal basis B for $\mathscr{B}|_T$

online: compute and threshold

$$\operatorname{dist}(w, \mathscr{B}) = \left\| (I - BB^\top) w \right\|$$

with noisy data w_{d}, the offline step is

- SVD truncation of $\mathscr{H}_T(w_{\mathrm{d}})$
- structured low-rank approximation of $\mathscr{H}_T(w_{\mathrm{d}})$
- model identification, using w_{d}

With unobserved excitation signal e, prior knowledge about e is needed

zero-mean white Gaussian (disturbance)

deterministic signal $\leadsto$ input estimation problem

the model describes $w_{\text{ext}} := \left[\begin{smallmatrix} e \\ w \end{smallmatrix}\right]$

- e — unobserved signal
- w — observed signal

Finding e is a linear least-norm problem

given a model $\mathscr{B}_{\text{ext}}$ that describes $w_{\text{ext}} := \left[\begin{smallmatrix} e \\ w \end{smallmatrix}\right]$

$$\widehat{e}_{\text{ln}} := \arg\min_{(\widehat{e},w) \in \mathscr{B}_{\text{ext}}|_T} \|\widehat{e}\|$$

exact recovery $\widehat{e}_{\text{ln}} = e$ is not possible

Deterministic input estimation is linear least-squares problem

Π_e / Π_w — projection of $w_{\text{ext}} := \left[\begin{smallmatrix} e \\ w \end{smallmatrix}\right]$ on e / w

given, $\widehat{\mathscr{B}}_{\text{ext}}|_T = \operatorname{image} B_{\text{ext}}$ (basis for $\widehat{\mathscr{B}}_{\text{ext}}|_T$)

$\widehat{e} := \Pi_e B_{\text{ext}} (\Pi_w B_{\text{ext}})^+ w$

Fault detection method with unobserved input

generalized distance measure:

$$\operatorname{dist}(w, \mathscr{B}_{\text{ext}}) := \min_{(\widehat{e}, \widehat{w}) \in \mathscr{B}|_T} \|w - \widehat{w}\|$$

offline: using $(e_{\text{d}}, w_{\text{d}})$, find basis B_{ext} for $\mathscr{B}_{\text{ext}}|_T$
and let $B_w := \Pi_w B_{\text{ext}}$

online: compute and threshold

$$\operatorname{dist}(w, \mathscr{B}_{\text{ext}}) = \left\|(I - B_w B_w^\top) w\right\|$$

Summary

assumptions:

- bounded complexity LTI system
- hyper-parameters: horizon T and lag ℓ
- different ways to deal with noise in offline data w_{d}

advantages:

- representation invariant distance measure
- can deal with unobserved disturbance signal
- cheap to compute online and simple to implement

other applications

References

IM, L. Huang, and F. Dörfler. *Data-driven control based on behavioral approach: From theory to applications in power systems. IEEE Control Systems Magazine, 43:28–68, 2023.*

interpolation and approximation of trajectories

IM and F. Dörfler. Data-driven dynamic interpolation and approximation. Automatica, 135:110008, 2022.

nonparametric frequency response estimation

IM and H. Ossareh. Finite-data nonparametric frequency response evaluation without leakage. Automatica, 159, 2024.

direct data-driven fault detection

IM, A. Muixí, S. Zlotnik, and P. Diez. A behavioral approach to direct data-driven fault detection. Mechanical Systems and Signal Processing, 245:113802, 2026.