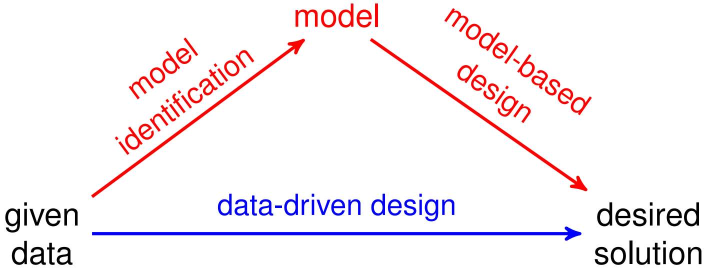
Data-driven dynamic interpolation and approximation

Ivan Markovsky
joint work with Florian Dörfler

CDC'19 Nice, France

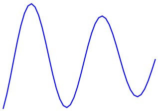
Data-driven = bypass model identification



Example: data-driven forecasting of sum-of-damped-exponentials signal

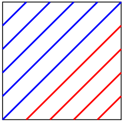
given: $\underbrace{y(1),\dots,y(t)}_{\text{"past" data}}$

find: $\underbrace{y(t+1),\dots,y(2t)}_{\text{"future" samples}}$



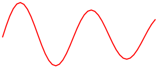
given data

$\mapsto$

rank  $\leq r$

matrix completion problem

$\mapsto$



prediction

$\rightsquigarrow$ Hankel structured low-rank completion

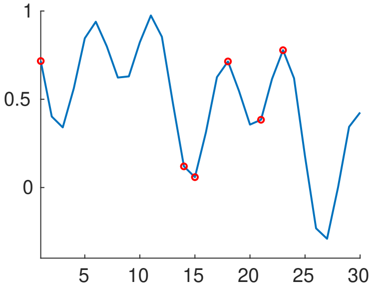
Simulation example of trajectory interpolation

```
%% parameters
T = 100; % number of data points
L = 30;  % interpolation horizon

%% interpolation points
t_given = randperm(L, n);
w_given = rand(n, 1);

%% data-driven interpolation
H = blkhank(wd, L);
g = H(t_given, :) \ w_given;
w = H * g;

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



Plan

Data-driven representation of LTI systems

Interpolation of trajectories

Generalizations and empirical validation

Outline

Data-driven representation of LTI systems

Interpolation of trajectories

Generalizations and empirical validation

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 $\quad :\Longleftrightarrow \quad$ $\mathscr{B}$ is subspace

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

$(\sigma w)(t) := w(t+1)$ — 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)

$\mathbf{l}(\mathscr{B})$ — 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 $\quad \widehat{\mathscr{B}} := \operatorname{span}\{\, w_{\mathrm{d}}, \sigma w_{\mathrm{d}}, \sigma^2 w_{\mathrm{d}}, \ldots \}$

fact: $\quad \widehat{\mathscr{B}} \in \mathscr{L}$ 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 $[1,L]$

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

fact: $\dim \mathscr{B}|_L = \mathbf{m}(\mathscr{B})L + \mathbf{n}(\mathscr{B})$, for all $L \geq \mathbf{l}(\mathscr{B})$

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

$$\mathscr{B}|_L = \mathscr{B}'|_L, \text{ for } L = \max\{\, \mathbf{l}(\mathscr{B}), \mathbf{l}(\mathscr{B}') \,\} + 1$$

Shifting and cutting w_d leads to Hankel matrix

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

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

define $\quad \widehat{\mathscr{B}}_L := \text{image}\, \mathscr{H}_L(w_\text{d})$

fact: $\quad \widehat{\mathscr{B}}_L \subseteq \mathscr{B}|_L$

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

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

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

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

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

$\widehat{\mathscr{B}}_L \subseteq \mathscr{B}|_L$, $L \geq \mathbf{l}(\mathscr{B})$, equality holds if and only if

$$\text{rank}\, \mathscr{H}_L(w_\text{d}) = L\mathbf{m}(\mathscr{B}) + \mathbf{n}(\mathscr{B})$$

sufficient conditions ("fundamental lemma"):

1. $w_\text{d} = \begin{bmatrix} u_\text{d} \\ y_\text{d} \end{bmatrix}$
2. $\mathscr{B}$ controllable
3. $\mathscr{H}_{L+\mathbf{n}(\mathscr{B})}(u_\text{d})$ 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 of trajectories

Generalizations and empirical validation

Data-driven interpolation is missing data recovery

given:

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

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

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

find:

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

Solution may not exist

A1: $\operatorname{rank} \mathscr{H}_L(w_\mathrm{d}) = \mathbf{m}(\mathscr{B})L + \mathbf{n}(\mathscr{B})$

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

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

($M|_{I_\text{given}}$ 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}

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

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

Data-driven interpolation method: solve system of linear equations

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

method:

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

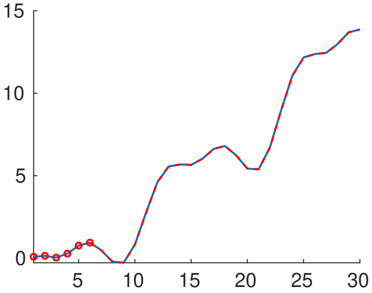
Simulation is special case of interpolation

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

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

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

%% results
plot(w), hold on,
plot(n:L, ws, 'r--')
plot(t_given, w_given, 'rO')
```



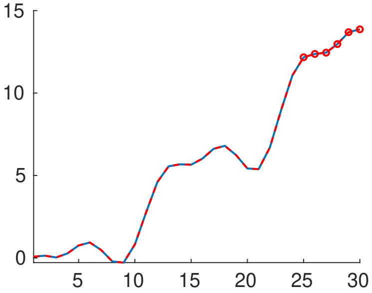
Simulation with “terminal conditions”

```
%% interpolation points
t_given = L-n+1:L;
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')
```



Outline

Data-driven representation of LTI systems

Interpolation of trajectories

Generalizations and empirical validation

approximation + missing data estimation

$$\begin{array}{ll} \text{minimize} & \text{over } g \text{ and } \widehat{w} \quad \big\| w|_{I_{\text{given}}} - \widehat{w}|_{I_{\text{given}}} \big\| \\ \text{subject to} & \widehat{w} = \mathscr{H}_L(w_{\text{d}})g \end{array}$$

data-driven filtering and control are special cases

interpolation + approximation + missing data

$\rightsquigarrow$ equality constrained least squares problem

multiple data trajectories $w_{\text{d}}^1, \ldots, w_{\text{d}}^N$

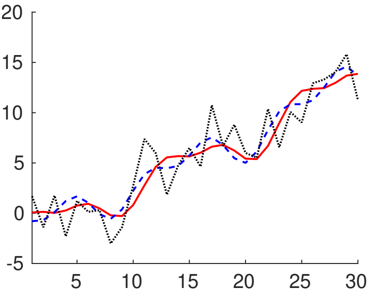
$$w = \begin{bmatrix} \mathscr{H}_L(w_{\text{d}}^1) & \cdots & \mathscr{H}_L(w_{\text{d}}^N) \end{bmatrix} g$$

Example: data-driven approximation (errors-in-variables Kalman smoothing)

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

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

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



efficient / real-time computation

non-parametric version of the Kalman filter

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, . . .

ℓ_1-norm regularization

in the noise-free case g can be chosen sparse

$$\|g\|_0 \leq L\mathbf{m}(\mathscr{B}) + \mathbf{n}(\mathscr{B})$$

impose sparsity in the case of noisy data

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

hyper-parameter: $\quad \lambda \quad \leftrightarrow \quad \mathbf{m}(\mathscr{B}), \mathbf{n}(\mathscr{B})$

Empirical validation on real-life datasets

	data set name	T	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

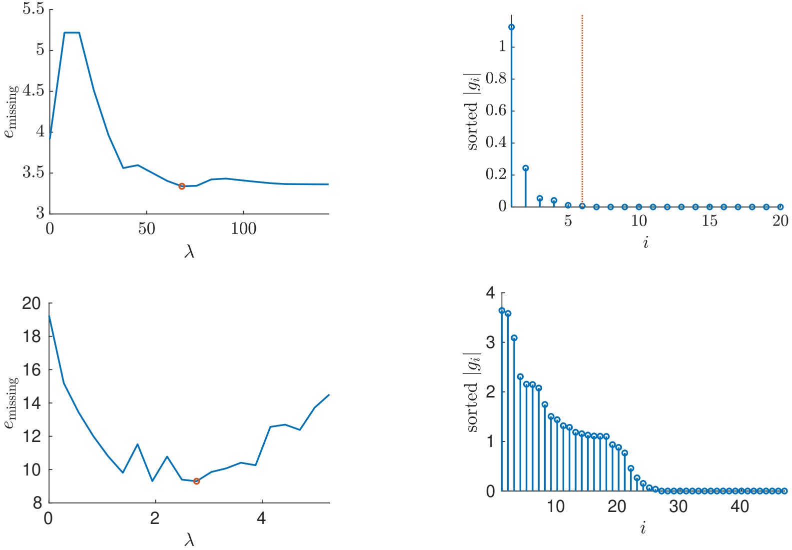
B. De Moor, et al.*DAISY: A database for identification of systems. Journal A, 38:4–5, 1997*

ℓ_1-norm regularization with optimized λ achieves the best performance

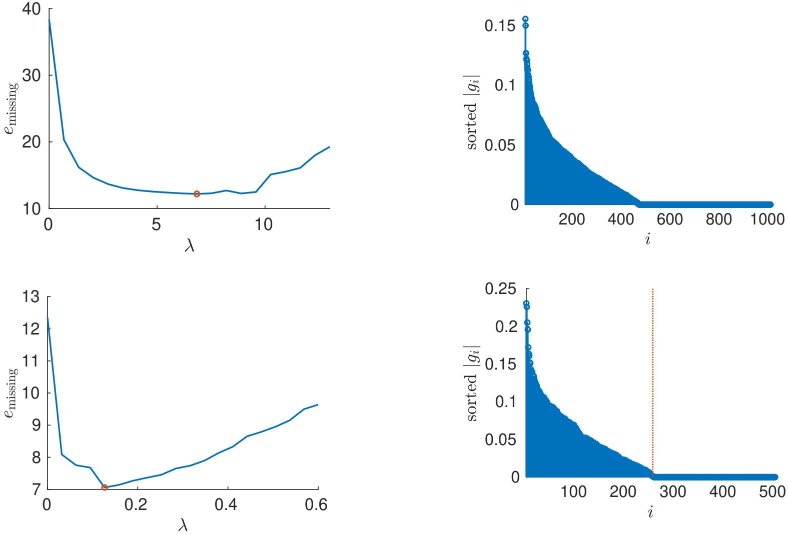
$$e_{\text{missing}} := \frac{\left\| w|_{I_{\text{missing}}} - \widehat{w}|_{I_{\text{missing}}} \right\|}{\left\| w|_{I_{\text{missing}}} \right\|} 100\%$$

	data set name	`pinv`	ML	ℓ_1-norm
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

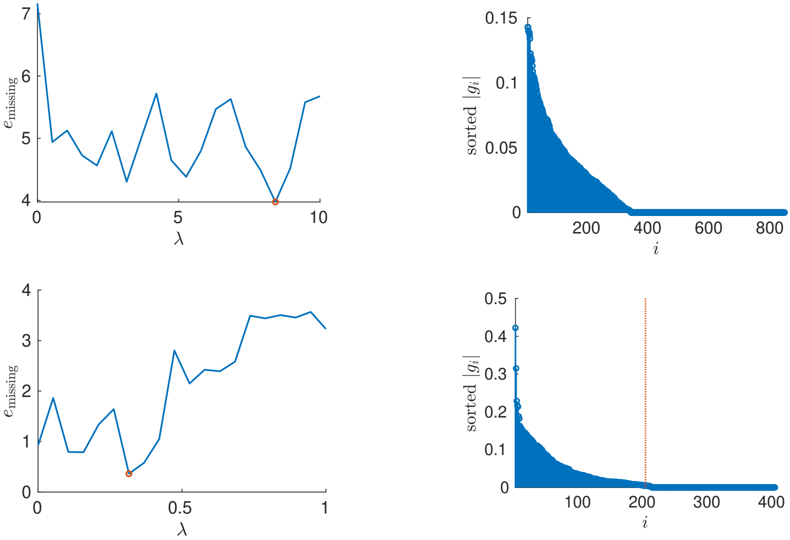
Tuning of λ and sparsity of g (datasets 1, 2)



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



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



System theory without transfer function and state space representations is possible

data-driven representation

leads to general, simple, practical methods

interpolation/approximation of trajectories

simulation, filtering and control are special cases
assumes only LTI dynamics; no hyper parameters

future work

efficient / real-time algorithms
sensitivity / statistical analysis
nonlinear systems