DYSCO course on low-rank approximation and its applications

Extensions and generalization

Ivan Markovsky

Vrije Universiteit Brussel

Vrije Universiteit Brussel

erc

Plan

Outline

Outline

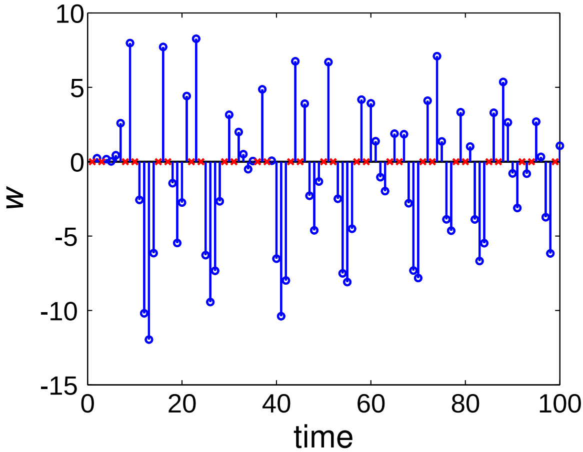
Why missing data?

- sensor failures
  measurements are accidentally corrupted
- compressive sensing
  measurements are intentionally skipped
- model-free signal processing
  missing data is what we aim to find

Exact identification with missing data

- the given data is exact
- data generating system is unknown but LTI
- problem is to interpolate the missing data
  (*cf.*, polynomial interpolation)
- special case: partial realization
  - given data — finite impulse response $h(1), \ldots, h(T)$
  - missing data — extension $h(T+1), \ldots$

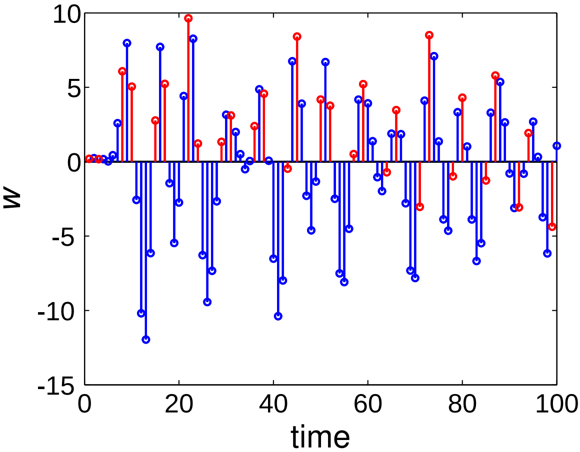
Example: exact SYSID with missing data



○ — 6th order autonomous LTI system's trajectory

× — missing data locations

Example: exact SYSID with missing data



○ — 6th order autonomous LTI system's trajectory

○ — interpolated data

The problem

- notation:

$\mathscr{I}_{\text{data}}$	—	given/specified elements of w
$w\vert_{\mathscr{I}_{\text{data}}}$	—	selects the elements $\mathscr{I}_{\text{data}}$ of w

- given: data $\mathscr{I}_{\text{data}}$ and $w|_{\mathscr{I}_{\text{data}}}$
- find: LTI system $\widehat{\mathscr{B}}$ of minimal order and $\widehat{w}$, such that

$$\widehat{w}|_{\mathscr{I}_{\text{data}}} = w|_{\mathscr{I}_{\text{data}}} \quad \text{and} \quad \widehat{w} \in \widehat{\mathscr{B}}$$

Equivalence to matrix completion

- the problem is equivalent to

$$\begin{array}{lll} \text{minimize} & \text{over } \widehat{w} & \operatorname{rank}\left(\mathscr{H}_L(\widehat{w})\right) \\ \text{subject to} & \widehat{w}|_{\mathscr{I}_{\text{data}}} = w|_{\mathscr{I}_{\text{data}}} & \end{array}$$

where

$$\mathscr{H}_L(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-L+1) \\ w(2) & w(3) & \cdots & w(T-L+2) \\ w(3) & w(4) & \cdots & w(T-L+3) \\ \vdots & \vdots & & \vdots \\ w(L) & w(L+1) & \cdots & w(T) \end{bmatrix}$$

- Hankel structured low-rank matrix completion

Special case: partial realization

- $\mathscr{I}_{\text{data}} = (1, \dots, T)$
- $w|_{\mathscr{I}_{\text{data}}} = \big(h(1), \dots, h(T)\big)$

$$\text{minimize over the ?'s} \quad \text{rank} \begin{bmatrix} h(1) & h(2) & h(3) & \cdots & h(T) \\ h(2) & h(3) & \iddots & h(T) & ? \\ h(3) & \iddots & \iddots & \iddots & ? \\ \vdots & h(T) & \iddots & \iddots & \vdots \\ h(T) & ? & ? & \cdots & ? \end{bmatrix}$$

Types of methods

- convex relaxations (nuclear norm heuristic)

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{w} \quad \|\mathscr{H}_L(\widehat{w})\|_* \\ \text{subject to} & \widehat{w}|_{\mathscr{I}_{\text{data}}} = w|_{\mathscr{I}_{\text{data}}} \end{array}$$

  replaces rank with the nuclear norm $\|\cdot\|_*$

- subspace methods

- local optimization based methods

Nuclear norm heuristic

- Hankel matrix nuclear norm minimization

$$\text{minimize} \quad \text{over } \widehat{w} \quad \|\mathscr{H}_L(\widehat{w})\|_*$$
$$\text{subject to} \quad \widehat{w}|_{\mathscr{I}_{\text{data}}} = w|_{\mathscr{I}_{\text{data}}}$$

- is a semidefinite optimization problem

$$\text{minimize} \quad \text{over } \widehat{w},\, U,\, V \quad \operatorname{trace}(U) + \operatorname{trace}(V)$$
$$\text{subject to} \quad \widehat{w}|_{\mathscr{I}_{\text{data}}} = w|_{\mathscr{I}_{\text{data}}}, \quad \begin{bmatrix} U & \mathscr{H}_L^\top(\widehat{w}) \\ \mathscr{H}_L(\widehat{w}) & V \end{bmatrix} \succeq 0$$

- $O(T^2)$ optimization variables $\quad$ (T — # of data points)

CVX code

```
function wh = hmc(w)

[T, q] = size(w); Idata = find(~isnan(w));
L = ceil((T + 1) / (q + 1));

cvx_begin sdp;
  variable wh(size(w));
  minimize norm_nuc(hankel(hh(1:L), hh(L:end)));
  subject to
    wh(Idata) == w(Idata);
cvx_end
```

Numerical example: partial realization

```
rand('seed', 0); r = 3; T = 10;
sys0 = drss(r);

h0 = impulse(sys0, 2 * T); h0 = h0(2:end);
h  = h0; h((T + 1):end) = NaN;

hh = hmc(h, T); err = norm(h0 - hh)
sv = svd(hankel(hh(1:T), hh(T:end)));
format long, first_sv = sv(1:(r + 1))
```

Output of CVX

```
Calling SDPT3: 210 variables, 91 equality constrain
----------------------------------------------------
...
number of iterations = 12
Total CPU time (secs) = 0.23
...

err =
9.250411145054003e-10

first_sv =
0.798479261343370
0.400697013978696
0.014660904007509
0.000000000297693
```

Subspace method by example

- order: $\ell = 2$, complete trajectory: $\bar{w}$
- $\implies R\mathscr{H}_3(\bar{w}) = 0$, for some $R \in \mathbb{R}^{1\times 3}$
- data: $w = (1, 2, \texttt{NaN}, 4, 5, \texttt{NaN}, 7, 8, \texttt{NaN}, 10, 11)$
- R can not be found from $\mathscr{H}_3(w)$

$$\begin{bmatrix} 1 & 2 & \texttt{NaN} & 4 & 5 & \texttt{NaN} & 7 & 8 & \texttt{NaN} \\ 2 & \texttt{NaN} & 4 & 5 & \texttt{NaN} & 7 & 8 & \texttt{NaN} & 10 \\ \texttt{NaN} & 4 & 5 & \texttt{NaN} & 7 & 8 & \texttt{NaN} & 10 & 11 \end{bmatrix}$$

- consider the matrix $\mathscr{H}_4(w)$

$$\begin{bmatrix} 1 & 2 & \mathtt{NaN} & 4 & 5 & \mathtt{NaN} & 7 & 8 \\ 2 & \mathtt{NaN} & 4 & 5 & \mathtt{NaN} & 7 & 8 & \mathtt{NaN} \\ \mathtt{NaN} & 4 & 5 & \mathtt{NaN} & 7 & 8 & \mathtt{NaN} & 10 \\ 4 & 5 & \mathtt{NaN} & 7 & 8 & \mathtt{NaN} & 10 & 11 \end{bmatrix}$$

- and select the columns in blue and red

$$\widetilde{H}^1 = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ \mathtt{NaN} & \mathtt{NaN} & \mathtt{NaN} \\ 4 & 7 & 10 \end{bmatrix} \quad \widetilde{H}^2 = \begin{bmatrix} 2 & 5 & 8 \\ \mathtt{NaN} & \mathtt{NaN} & \mathtt{NaN} \\ 4 & 7 & 10 \\ 5 & 8 & 11 \end{bmatrix}$$

- removing the rows of NaN's

$$\underbrace{\begin{bmatrix} 1 & -3/2 & 1/2 \end{bmatrix}}_{R^1} \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 4 & 7 & 10 \end{bmatrix} = 0 \quad \underbrace{\begin{bmatrix} 1 & -3 & 2 \end{bmatrix}}_{R^2} \begin{bmatrix} 2 & 5 & 8 \\ 4 & 7 & 10 \\ 5 & 8 & 11 \end{bmatrix} = 0$$

- we have

$$\underbrace{\begin{bmatrix}1 & -3/2 & 0 & 1/2\end{bmatrix}}_{\widetilde{R}^1}\widetilde{H}^1 = 0, \quad \underbrace{\begin{bmatrix}1 & 0 & -3 & 2\end{bmatrix}}_{\widetilde{R}^2}\widetilde{H}^2 = 0$$

- by construction $\begin{bmatrix}\widetilde{R}^1 \\ \widetilde{R}^2\end{bmatrix}\mathscr{H}_4(\bar{w}) = 0$, so that

$$\widetilde{R}(z) = \begin{bmatrix}\widetilde{R}^1(z) \\ \widetilde{R}^2(z)\end{bmatrix} = \begin{bmatrix}z^0 - 3/2z^1 + 1/2z^3 \\ z^0 - 3z^2 + 2z^3\end{bmatrix}$$

is a (nonminimal) kernel repr. of the system

- a minimal representation is given by

$$R(z) := \mathrm{GCD}\big(\widetilde{R}^1(z), \widetilde{R}^2(z)\big) = z^0 - 2z^1 + z^2$$

- once R is computed, it is trivial to complete the data

$$\begin{aligned}\bar{w} &= (1 \quad 2 \quad \texttt{NaN} \quad 4 \quad 5 \quad \texttt{NaN} \quad 7 \quad 8 \quad \texttt{NaN} \quad 10 \quad 11)\\ \widehat{w} &= (1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11)\end{aligned}$$

Current/future work

- I. Markovsky. Exact identification with missing data. In *Proc. of the 52nd IEEE Conference on Decision and Control*, pages 151–155, Florence, Italy, 2013
- generalization to MIMO systems $\rightsquigarrow$ $O(T)$ method
- reduction to minimal representation
- in case of noisy data, it is model reduction
- possible approach: approximate common divisor

Exact identification with missing data

- the problem is equivalent to finding $\widehat{w}$, such that

$$\underbrace{\| w|_{\mathscr{I}_\text{g}} - \widehat{w}|_{\mathscr{I}_\text{g}} \| = 0}_{\text{exact data}} \quad \text{and} \quad \underbrace{\operatorname{rank}\big(\mathscr{H}_L(\widehat{w})\big) \leq r}_{\text{of an LTI system}}$$

where r is bound on the model complexity and

$$\mathscr{H}_L(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-L+1) \\ w(2) & w(3) & \cdots & w(T-L+2) \\ w(3) & w(4) & \cdots & w(T-L+3) \\ \vdots & \vdots & & \vdots \\ w(L) & w(L+1) & \cdots & w(T) \end{bmatrix}$$

- Hankel structured low-rank matrix completion

Approx. identification with missing data

- given w and r

$$\text{minimize} \quad \text{over } \widehat{w} \quad \underbrace{\| w|_{\mathscr{I}_\text{g}} - \widehat{w}|_{\mathscr{I}_\text{g}} \|}_{\text{approximation error}}$$

$$\text{subject to} \quad \underbrace{\operatorname{rank}\big(\mathscr{H}_L(\widehat{w})\big) \leq r}_{\substack{\widehat{w} \text{ is trajectory of} \\ \text{bounded complexity LTI system}}}$$

- approx. Hankel structured low-rank matrix completion

Main idea

- element-wise nonnegative weights $w_i(t) \leftrightarrow v_i(t)$
- weighted cost function

$$\|w - \widehat{w}\|_v := \sqrt{\sum_{t=1}^{T} \sum_{i=1}^{q} v_i(t)\big(w_i(t) - \widehat{w}_i(t)\big)^2}$$

- zero weight $v_i(t) = 0 \leftrightarrow$ missing value $w_i(t)$
- $v_i(t) = \dfrac{1}{\text{"variance of the noise on } w_i(t)\text{"}}$
- zero weight $\leftrightarrow$ infinite noise variance

Problem

- with $v_i(t) = \begin{cases} 1, & \text{if } w_i(t) \text{ is given} \\ 0, & \text{if } w_i(t) \text{ is missing} \end{cases}$

$$\|w|_{\mathscr{I}_\text{g}} - \widehat{w}|_{\mathscr{I}_\text{g}}\| = \|w - \widehat{w}\|_v$$

- and the problem is

$$\text{minimize} \quad \text{over } \widehat{w} \quad \|w - \widehat{w}\|_v \qquad \text{(SLRA)}$$
$$\text{subject to} \quad \operatorname{rank}\big(\mathscr{H}_L(\widehat{w})\big) \leq r$$

- weighted Hankel structured low-rank approximation

Nuclear norm heruistic

- replacing rank with $\|\cdot\|_*$, the problem is relaxed to

$$\text{minimize} \quad \|W_\mathrm{l}\mathscr{H}(\widehat{w})W_\mathrm{r}\|_* + \lambda\|w - \widehat{w}\|_2^2$$

- fast solution method:

  Z. Liu, A. Hansson, and L. Vandenberghe. Nuclear norm system identification with missing inputs and outputs. *Control Lett.*, 62:605–612, 2013

Parameter optimization

- using the kernel parameterization

$$\operatorname{rank}\big(\mathscr{H}_L(\widehat{w})\big) \leq r \iff \begin{array}{c} R\mathscr{H}_L(\widehat{w}) = 0 \\ R \in \mathbb{R}^{\mathtt{p} \times qL} \text{ full row rank (f.r.r.)} \end{array}$$

q	—	# of variables
$\mathtt{p} := qL - r$	—	co-rank (rank deficiency)

- (SLRA) becomes

$$\begin{array}{lll} \text{minimize} & \text{over } \widehat{w} \text{ and } R & \|w - \widehat{w}\|_v \\ \text{subject to} & R\mathscr{S}(\widehat{p}) = 0 \text{ and } R \text{ f.r.r.} & \end{array} \qquad (\text{SLRA}_R)$$

VARPRO-like solution method

- (SLRA_R) is separable in $\widehat{p}$ and R, *i.e.*,

$$\text{minimize} \quad \text{over f.r.r. } R \in \mathbb{R}^{\mathrm{p} \times qL} \quad f(R) \qquad \text{(OUTER)}$$

  where

$$f(R) := \min_{\widehat{w}} \; \|w - \widehat{w}\|_v \text{ s.t. } R\mathscr{H}_L(\widehat{w}) = 0 \qquad \text{(INNER)}$$

- (INNER) is a (generalized) least norm problem
- $\widehat{p}$ is eliminated (projected out) of (SLRA_R)

Evaluation of $f(R)$ with missing data

$$f = \min_{x,y} \quad x^\top x \quad \text{subject to} \quad Ax + By = c \qquad \text{(GLN)}$$

Lemma under the following assumptions

A1. B is full column rank

A2. $1 \leq \dim(c) - \dim(y) \leq \dim(x)$

A3. $\bar{A} := B^\perp A$ is full row rank

(GLN) has a unique solution

$$f = c^\top (B^\perp)^\top (\bar{A}\bar{A}^\top)^{-1} B^\perp c,$$
$$x = \bar{A}^\top (\bar{A}\bar{A}^\top)^{-1} B^\perp c, \quad y = B^+(c - Ax)$$

Proof

Under A1 and A2, $\text{rank}(B) = n_y$ and

$$TB = \begin{bmatrix} B^+ \\ B^\perp \end{bmatrix} B = \begin{bmatrix} T^+ B \\ T^\perp B \end{bmatrix} = \begin{bmatrix} I_{n_y} \\ 0 \end{bmatrix}, \qquad \det(T) \neq 0$$

Then

$$\begin{bmatrix} B^+ Ax \\ B^\perp Ax \end{bmatrix} + \begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} B^+ c \\ B^\perp c \end{bmatrix}.$$

The first equation

$$y = B^+(c - Ax)$$

uniquely determines y, given x. The second equation

$$B^\perp Ax = B^\perp c \qquad (*)$$

defines a linear constraint for x only.

Proof

By assumption A2, $(*)$ is an underdetermined system of linear equations. Therefore, (GLN) is equivalent to the following standard weighted least norm problem

$$f = \min_{x} \; x^\top x \quad \text{subject to} \quad B^\perp A x = B^\perp c. \qquad \text{(GLN')}$$

By assumption A3 the solution is unique.

About the assumptions

- A1 and A3 ensure uniqueness of y
- otherwise, $y \in B^{+}(c - Ax) + \text{null}(B)$
- A2 ensures feasibility with a nontrivial solution
- with $m = n_y$, (GLN) has trivial solution $f = 0$
- with $m - n_y > n_x$, (GLN) generically has no solution

Examples

1. autonomous system identification with missing data
   - 2nd order, $T = 50$, $y = \bar{y} +$ white noise
   - periodically missing data with period 3

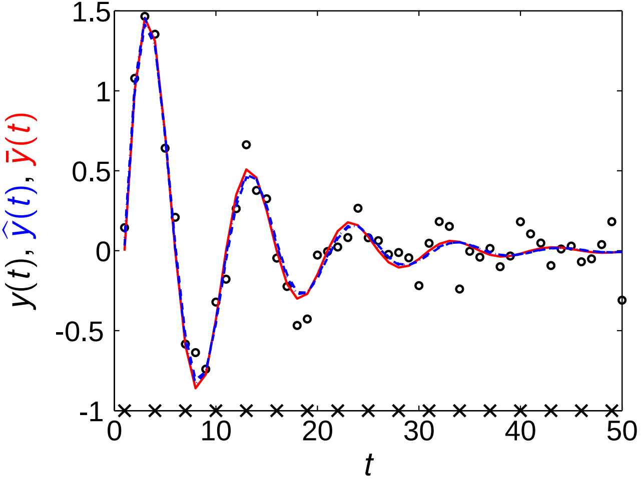
2. data-driven simulation (as missing data estimation)

3. data-driven control (as missing data estimation)

- I. Markovsky and K. Usevich. Structured low-rank approximation with missing data. *SIAM J. Matrix Anal. Appl.*, pages 814–830, 2013

- I. Markovsky. Approximate identification with missing data. In *Proc. of the 52nd IEEE Conference on Decision and Control*, pages 156–161, Florence, Italy, December 2013

Autonomous system identification



true data, ○ — noisy data × — missing approx.

Model-free simulation

- second order SISO system, defined by

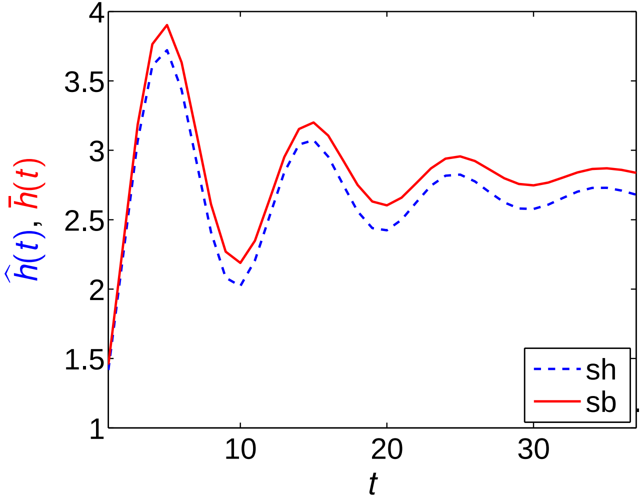
$$y(t) = 1.456y(t-1) - 0.81y(t-2) + u(t) - u(t-1)$$

- w^1 is noisy trajectory generated from random input

- w^2 is the impulse response estimate $\bar{h}$, *i.e.*,

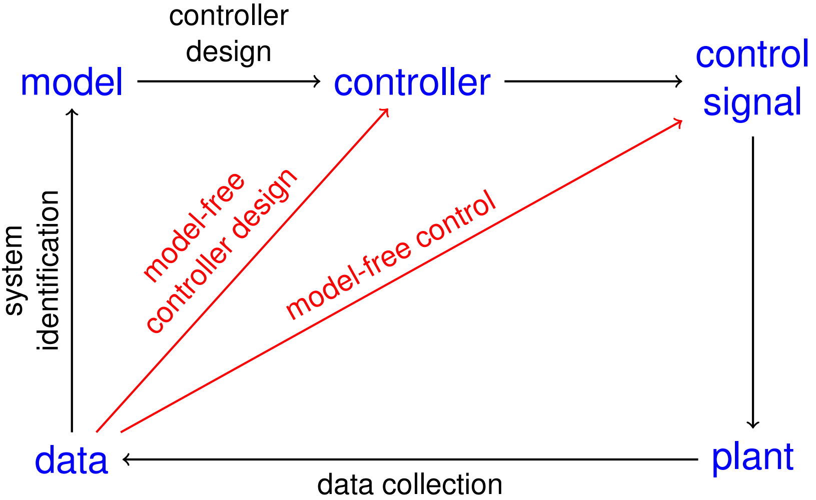
$$u^2 = (\underbrace{0, \ldots, 0}_{\ell}, \underbrace{1, 0, \ldots, 0}_{\text{pulse input}})$$

$$y^2 = (\underbrace{0, \ldots, 0}_{\ell}, \underbrace{\widehat{h}(0), \widehat{h}(1), \ldots, \widehat{h}(T_2 - \ell - 1)}_{\text{impulse response — missing data}})$$



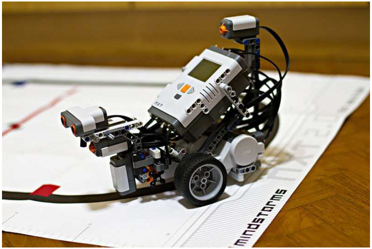
true impulse response $\bar{h}$ model free estimate $\widehat{h}$

Model-free control



Classical LTI optimal tracking control

- given:
  - system $\mathscr{B} \in \mathscr{L}$
  - desired output y_r



- find: input u, such that

$$\min_{(u,y)\in\mathscr{B}} \|y_\mathrm{r} - y\|$$

- there are different algorithms to solve the problem (Riccati equation, spectral factorization, . . .)

- they depend on the representation of the model $\mathscr{B}$ (state-space, transfer function, . . .)

Model-free LTI optimal tracking control

- given:
  - trajectory w^1 (of a system $\bar{\mathscr{B}} \in \mathscr{L}_{\mathrm{m},\ell}$)
  - desired output $y_\mathrm{r} = y^2$
- find: input $\widehat{u}^2$, such that

$$\begin{array}{ll} \text{minimize} & \underbrace{\|w^1 - \widehat{w}^1\|^2}_{\substack{\text{misfit}\\\text{error}}} + \underbrace{\|y_\mathrm{r} - \widehat{y}^2\|^2}_{\substack{\text{tracking}\\\text{error}}} \\ \text{subject to} & \widehat{w}^1, \widehat{w}^2 \in \widehat{\mathscr{B}} \in \mathscr{L}_{\mathrm{m},\ell} \end{array} \qquad \text{(MFT)}$$

- $\widehat{\mathscr{B}}$ in (MFT) is needed to define the problem
- in a model-free method, $\widehat{\mathscr{B}}$ is not identified explicitly

Solution by SLRA with missing data

- (MFT) is equivalent to

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{w} \text{ and } \widehat{\mathscr{B}} \quad \|w - \widehat{w}\|_\alpha \\ \text{subject to} & \widehat{w} \subset \widehat{\mathscr{B}} \in \mathscr{L}_{\mathrm{m},\ell} \end{array}$$

  the control input u^2 is missing data

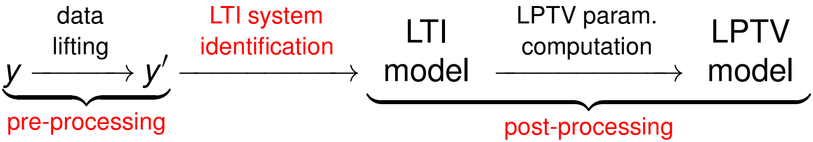
- this leads to mosaic-Hankel SLRA with missing data

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{w} \quad \|w - \widehat{w}\|_\alpha \\ \text{subject to} & \operatorname{rank}\big(\mathscr{H}_{\ell+1}(\widehat{w})\big) \leq q\ell + \mathrm{m} \end{array}$$

- the only truly model free solution methods I know of are based on the nuclear norm heuristic

Outline

Outline



1. $O(T\mathrm{p}L^2)$ algorithm for LPTV system realization
   - T — # of samples
   - p — # of outputs (dim(y))
   - L — upper bound of the order

2. algorithm for LPTV maximum likelihood identification
   I. Markovsky, J. Goos, K. Usevich, and R. Pintelon.
   Realization and identification of autonomous linear periodically time-varying systems. *Automatica*, 2014

Autonomous LPTV systems

- state space representation (σ — shift operator)

$$\mathscr{B}(A,C) := \{\, y \mid \sigma x = Ax,\ y = Cx,\ x(1) = x_{\text{ini}} \in \mathbb{R}^n \,\}$$

A and C are functions of time

- change of basis, *i.e.*,

$$\mathscr{B} = \mathscr{B}(A,C) = \mathscr{B}(\widehat{A},\widehat{C}), \qquad \begin{aligned} \widehat{A} &= \sigma V A V^{-1} \\ \widehat{C} &= C V^{-1} \end{aligned}$$

- P-periodicity

$$A = \sigma^P A, \qquad C = \sigma^P C, \qquad V = \sigma^P V$$

Problem formulation

realization

- given: $y = \big(y(1), \dots, y(T)\big)$, period P, and order $\mathtt{n}$
- find: $\widehat{A}$, $\widehat{C}$, such that $y \in \mathscr{B}(\widehat{A}, \widehat{C})$

identification

- given: $y = \big(y(1), \dots, y(T)\big)$, period P, and order $\mathtt{n}$

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{y} \text{ and } \widehat{\mathscr{B}} \quad \|y - \widehat{y}\|_2 \\ &\text{subject to} \quad \widehat{y} \in \widehat{\mathscr{B}} \in \mathscr{L}_{0,\mathtt{n},P} \end{aligned}$$

$\mathscr{L}_{0,\mathtt{n},P}$ — set of autonomous LPTV systems with order at most $\mathtt{n}$ and period P $(0 =$ no inputs$)$

"lifting" operator

$$\big(y(1),\dots,y(T)\big) = y \mapsto y' = \big(y'(1),\dots,y'(T')\big),\ T' := \lfloor T/P \rfloor$$

$$y' = \mathrm{lift}_P(y) = \left(\begin{bmatrix} y(1) \\ \vdots \\ y(P) \end{bmatrix}, \begin{bmatrix} y(P+1) \\ \vdots \\ y(2P) \end{bmatrix}, \dots, \begin{bmatrix} y((T'-1)P) \\ \vdots \\ y(T'P) \end{bmatrix} \right)$$

Theorem 1

- $\mathscr{B}(A,C)$ — LPTV of order $\mathtt{n}$, period P, with $\mathtt{p}$ outputs
- $\mathrm{lift}_P\big(\mathscr{B}(A,C)\big)$ — LTI of order $\mathtt{n}$, with $\mathtt{p}' := \mathtt{p}P$ outputs

Notation

- block-Hankel matrix

$$\mathscr{H}_L(y) := \begin{bmatrix} y(1) & y(2) & y(3) & \cdots & y(T-L+1) \\ y(2) & y(3) & \iddots & & y(T-L+1) \\ y(3) & \iddots & & & \vdots \\ \vdots & & & & \\ y(L) & y(L+1) & \cdots & & y(T) \end{bmatrix}$$

- extended observability matrix

$$\mathscr{O}_L(A,C) := \begin{bmatrix} C(1) \\ C(2)A(1) \\ C(3)A(2)A(1) \\ \vdots \\ C(L)A(L-1)A(L-2)\cdots A(1) \end{bmatrix}$$

Identification of the lifted system

- $\text{lift}_P\big(\mathscr{B}(A,C)\big)$ admits $\mathtt{n}$th order repr. $\mathscr{B}(\Phi,\Psi)$
- $y' \mapsto (\widehat{\Phi},\widehat{\Psi})$ is classical realization problem
- can be solved, *e.g.*, by Kung's method

$$\underbrace{\mathscr{H}_L(y')}_{\text{Hankel}} = \underbrace{\mathscr{O}(\widehat{\Phi},\widehat{\Psi})}_{\mathbf{O}}\underbrace{\mathscr{O}(\widehat{\Phi}^\top,\widehat{x}_{\text{ini}}^\top)}_{\mathbf{C}} \qquad \begin{array}{l} \mathbf{O} \in \mathbb{R}^{L\mathtt{p}'\times\mathtt{n}} \\ \mathbf{C} \in \mathbb{R}^{\mathtt{n}\times(T'-L)} \end{array}$$

- $\widehat{\Phi}^\top$ is a solution of the shift equation

$$\underline{\mathbf{O}}\widehat{\Phi} = \overline{\mathbf{O}}$$

- $\widehat{\Psi}$ is the first block-element of $\mathbf{O}$

Computation of the model parameters

Theorem 2

- define $\widehat{A} = \big(\widehat{A}_1, \ldots, \widehat{A}_P\big)$ and $\widehat{C} = \big(\widehat{A}_1, \ldots, \widehat{A}_P\big)$ via

$$\widehat{A}_1 = \ldots = \widehat{A}_{P-1} = I_n, \quad \widehat{A}_P := \widehat{\Phi}$$

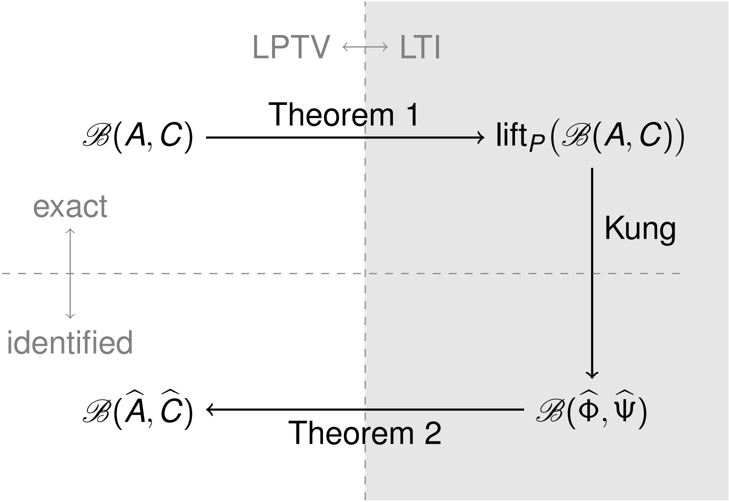
$$\mathrm{col}(\widehat{C}_1, \ldots, \widehat{C}_P) := \widehat{\Psi}, \quad \widehat{C}_i \in \mathbb{R}^{\mathrm{p} \times \mathrm{n}}$$

$$(\text{note that } \widehat{\Psi} = \widehat{A}_P \widehat{A}_{P-1} \cdots \widehat{A}_2 \widehat{A}_1)$$

- $\mathscr{B}(\widehat{\Phi}, \widehat{\Psi})$ (LTI) is equivalent to $\mathscr{B}(\widehat{A}, \widehat{C})$ (LPTV), *i.e.*,

$$\mathscr{B}(\widehat{\Phi}, \widehat{\Psi}) = \mathrm{lift}_P\big(\mathscr{B}(\widehat{A}, \widehat{C})\big)$$

Summary



LPTV ⟷ LTI
$\mathscr{B}(A, C)$
Theorem 1
$\text{lift}_P\big(\mathscr{B}(A, C)\big)$
exact
Kung
identified
$\mathscr{B}(\widehat{A}, \widehat{C})$
Theorem 2
$\mathscr{B}(\widehat{\Phi}, \widehat{\Psi})$

Modified method

- using the “transposed” lifted sequence

$$y'^{\top} := \big(y'^{\top}(1), \dots, y'^{\top}(T')\big), \qquad y'^{\top}(t) \in \mathbb{R}^{1\times \mathrm{p}'}$$

- the Hankel matrix factorization becomes

$$\underbrace{\mathscr{H}_L(y'^{\top})}_{\text{Hankel}} = \underbrace{\mathscr{O}_L(\widehat{\Phi}^{\top}, x_{\text{ini}}^{\top})}_{\mathbf{O}} \cdot \underbrace{\mathscr{O}_{T'-L+1}^{\top}(\widehat{\Phi}, \widehat{\Psi})}_{\mathbf{C}} \qquad \begin{array}{l} \mathbf{O} \in \mathbb{R}^{L\times \mathrm{n}} \\ \mathbf{C} \in \mathbb{R}^{\mathrm{n}\times \mathrm{p}(T-L)} \end{array}$$

- $\widehat{\Phi}^{\top}$ is a solution of the shift equation

$$\underline{\mathbf{O}}\widehat{\Phi}^{\top} = \overline{\mathbf{O}}$$

- $\widehat{\Psi}^{\top}$ is the first block element of **C**

Identification

- the SYSID problem is equivalent to

$$\text{minimize} \quad \text{over } \widehat{y} \quad \|y - \widehat{y}\|_2 \qquad \text{subject to} \quad \operatorname{rank}\left(\mathscr{H}_{\mathrm{n}+1}\left(\operatorname{lift}_P(\widehat{y}^\top)\right)\right) \leq \mathrm{n} \tag{SLRA}$$

- (SLRA) is Hankel structured low-rank approximation

- existing methods can be used; we use VARPRO method, based on the kernel representation

$$\operatorname{rank}\left(\mathscr{H}_{\mathrm{n}+1}\left(\operatorname{lift}_P(\widehat{y}^\top)\right)\right) \leq \mathrm{n} \quad \iff$$
$$\exists\, R^{1\times(\mathrm{n}+1)}, \quad R\mathscr{H}_{\mathrm{n}+1}\left(\operatorname{lift}_P(\widehat{y}^\top)\right) = 0, \quad RR^\top = 1$$

Simulation setup

- the data is generated by an output error model

  $y = \bar{y} + \widetilde{y}, \quad \text{where} \quad \bar{y} \in \bar{\mathscr{B}} \in \mathscr{L}_{0,\mathrm{n},P}$ and $\widetilde{y} \sim \mathrm{N}(0, s^2 I_{\mathrm{p}})$

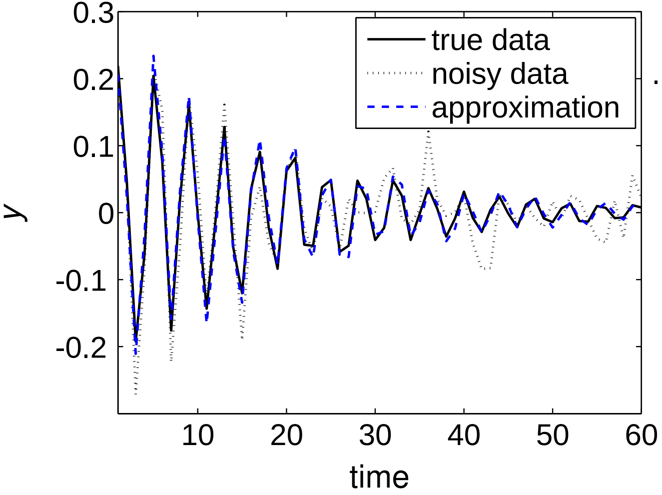
  split into identification (3/4) and validation (1/4) parts

- $\bar{\mathscr{B}}$ is Mathieu oscillator—spring-mass-damper system with time-periodic spring stiffness

$$\bar{A}_\tau = \begin{bmatrix} 0 & 1 \\ \bar{a}_1 & \bar{a}_{2,\tau} \end{bmatrix}, \quad \bar{C}_\tau = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

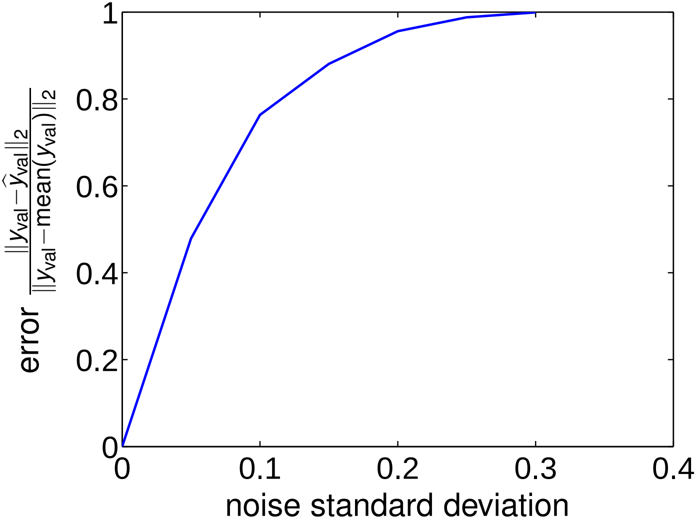
- in the example $P = 3$ and $T' = 20$

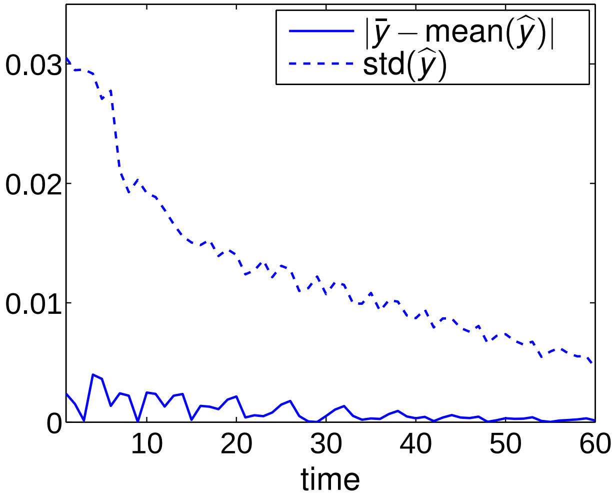
The data and its approximation



.

Average approximation error





$|\bar{y} - \text{mean}(\widehat{y})|$
$\text{std}(\widehat{y})$
0.03
0.02
0.01
0
10
20
30
40
50
60
time

Outline

Joint work with K. Usevich

Problem formulation

- Grassmann manifold

$$\mathscr{G}_{\mathtt{m}}^{q} := \{\, \mathscr{B} \mid \mathscr{B} \text{ is } \mathtt{m}\text{-dim. subspace of } \mathbb{R}^{q} \,\}$$

  the set of static linear models with complexity $\mathtt{m}$

- $\mathscr{L}_{\mathtt{m},0}^{q}$ — static linear models with bounded complexity

- optimization on a manifold

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{G}_{\mathtt{m}}^{q} \quad f(\mathscr{B}) \qquad \text{(GO)}$$

- $f(\mathscr{B})$ can be error of fit between data $\mathscr{D}$ and model $\mathscr{B}$

Parameterizations

- optimize over a basis P of $\mathscr{B}$, $\mathscr{B} = \text{image}(P)$

$$\text{minimize} \quad \text{over f.c.r. } P \in \mathbb{R}^{q \times \mathtt{m}} \quad f\big(\text{image}(P)\big)$$

- optimize over a basis R of $\mathscr{B}^{\perp}$, $\mathscr{B} = \text{image}(R^{\perp})$

$$\text{minimize} \quad \text{over f.r.r. } R \in \mathbb{R}^{(q-\mathtt{m}) \times q} \quad f\big(\text{image}(R^{\perp})\big)$$

- how to impose the "f.c.r." / "f.r.r." constraints?
- related issue: nonuniqueness of the basis

  $\mathscr{B} = \text{image}(P) = \text{image}(PU)$, for all nonsingular $U \in \mathbb{R}^{\mathtt{m} \times \mathtt{m}}$

- there is a smaller number of "effective parameters"

Solution approaches

- w.l.g. take P and R orthonormal:

$$P^\top P = I \qquad \text{and} \qquad RR^\top = I_{q-\mathrm{m}}$$

$\rightsquigarrow$ optimization with *quadratic equality constraints*

- fix a full rank block, *e.g.*, I:

$$P = \begin{bmatrix} I \\ X \end{bmatrix} \qquad \text{and} \qquad R = \begin{bmatrix} I & Y \end{bmatrix}$$

$X \in \mathbb{R}^{(q-\mathrm{m})\times \mathrm{m}}$ and $Y \in \mathbb{R}^{\mathrm{m}\times(q-\mathrm{m})}$ are free variables

- covers almost all subspaces, but not the whole $\mathscr{G}_{\mathrm{m}}^{q}$

Switching permutations

- introduce a permutation matrix Π, such that

$$P = \Pi \begin{bmatrix} I \\ X \end{bmatrix} \qquad \text{and} \qquad R = \begin{bmatrix} I & Y \end{bmatrix} \Pi$$

- the optimization is then over X / Y and Π
- due to choice of Π, this is combinatorial problem
- in practice *any* fix permutation is almost always sufficient to solve the problem
- $\|X\|$ may be large at the solution

Practical algorithm

- monitor $\|X\|$ throughout the iterations
- select adaptively different permutations
- details: K. Usevich and I. Markovsky. Optimization on a Grassmann manifold with application to system identification. *Automatica*, 2014
- implementation:

  http://slra.github.io/software.html

Penalty method for orthonormal bases

Theorem

For any $\gamma > 0$*, the local minima of*

$$\text{minimize} \quad \text{over } R \in \mathbb{R}^{(q-\mathtt{m})\times q} \quad f\big(\ker(R)\big) \quad \text{s.t.} \quad RR^\top = I_{q-\mathtt{m}}$$

coincide with the local minimal of

$$\text{minimize} \quad \text{over } R \in \mathbb{R}^{(q-\mathtt{m})\times q} \quad f\big(\ker(R)\big) + \gamma \|RR^\top - I_{q-\mathtt{m}}\|_F^2.$$

- details and implementation:

  `http://slra.github.io/software.html`

Other methods in system identification

- local coordinates

  T. McKelvey, A. Helmersson, and T. Ribarits. Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica*, 40:1629–1635, 2004

- pseudo-inverse of the Jacobian matrix

  R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, Piscataway, NJ, second edition, 2012

Outline

Joint work with M. Ishteva

Linearly structured matrices

$$D = \mathscr{S}(p) = p_1 S_1 + \cdots + p_{n_p} S_{n_p}$$

- image representation $D = PL$
- how to impose the structure via the factors P, L?
- for Hankel matrices,

  P, L have Vandermonde structure
- for general $\mathscr{S}$,

  the structure of P and L is an open problem

Penalty method

- orthogonal projection on image($\mathscr{S}$)

$$\Pi_{\mathscr{S}}(D) = \mathscr{S}\big(\Pi_{\mathbf{S}} \operatorname{vec}(D)\big)$$

where $\mathbf{S} = \begin{bmatrix} \operatorname{vec}(S_1) & \cdots & \operatorname{vec}(S_{n_p}) \end{bmatrix}$

- penalty method

$$\min_{P,L} \|D - \Pi_{\mathscr{S}}(PL)\| + \lambda \|PL - \Pi_{\mathscr{S}}(PL)\|$$

  - $\|D - \Pi_{\mathscr{S}}(PL)\|$ — approximation error term
  - $\|PL - \Pi_{\mathscr{S}}(PL)\|$ — distance to structured matrix

- for $\lambda = \infty$ the penalized problem recovers the structured low-rank solution

Alternating projections method

- solve iteratively
  - $\min_L \|p - \Pi_{\mathbf{S}} \operatorname{vec}(PL)\| + \lambda \|PL - \Pi_{\mathscr{S}}(PL)\|$
  - $\min_P \|p - \Pi_{\mathbf{S}} \operatorname{vec}(PL)\| + \lambda \|PL - \Pi_{\mathscr{S}}(PL)\|$

- each problem is a linear least squares problem

- start with small λ and initial L, P obtained from the unstructured LRA

- increase λ in the course of the optimization (homotopy method)

- M. Ishteva, K. Usevich, and I. Markovsky. Regularized structured low-rank approximation. Technical report, Vrije Univ. Brussel, 2013. Submitted on 02/08/2013 to *SIAM J. Matrix Anal. Appl.*

Outline

Model reduction methods

- Balanced model reduction

  B. Moore. Principal component analysis in linear systems: Controllability, observability and model reduction. *IEEE Trans. Automat. Control*, 26(1):17–31, 1981

- Proper orthogonal decomposition:

  heruistic method for nonlinear state space models

$$\begin{aligned} \sigma x &= f(x, u) \\ y &= h(x, u) \end{aligned}$$

  with potentially very large $\dim(x) = \mathtt{n}$

Proper orthogonal decomposition

- approximate the matrix of state "snapshots"

$$X := \begin{bmatrix} x(t_1) & \cdots & x(t_N) \end{bmatrix} \in \mathbb{R}^{\mathrm{n} \times N}$$

  by a rank-$r \ll \mathrm{n}$ matrix, using, *e.g.*, the truncated SVD

$$X = U \Sigma V^\top \approx U_r \Sigma_r V_r^\top$$

- reduced state: $\widehat{x} = U_r^\top x$, evolves in R^r
- reduced order system:

$$\sigma \widehat{x} = U_r^\top f(U_r \widehat{x}, u), \quad y = h(U_r \widehat{x}, u)$$

- issues: choice of snapshots, error bounds, ...

Low-rank approximation at MTNS 2014

- Mathematical Theory of Networks and Systems

  http://fwn06.housing.rug.nl/mtns2014

- mini-course with lectures by
  - M. Isheva on the "factorization approach"
  - K. Usevich on "Grassmann manifold minimization"

- invited sessions (under review!) on
  - matrix problems/methods for LRA
  - tensor problems/methods for LRA

Reproducible research

> *“An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.”*

- J. Buckheit and D. Donoho. *Wavelets and statistics*, chapter "Wavelab and reproducible research". Springer-Verlag, Berlin, New York, 1995
  - Can others (easily) redo your simulations?
  - Can *you* (easily) redo the simulations?
- Tools: org-mode (emacs), `publish` (matlab), and git

Literate programming

"At first, I thought programming was primarily analogous to musical composition—to the creation of intricate patterns, which are meant to be performed. But lately I have come to realize that a far better analogy is available: Programming is best regarded as the process of creating works of literature, which are meant to be read."

- D. Knuth. *Literate programming*. Cambridge University Press, 1992

- Tools: noweb (any language), sweave (R), . . .

  N. Ramsey. Literate programming simplified. *IEEE Software*, 11:97–105, 1994