Approximate system identification with missing data

Ivan Markovsky

Vrije Universiteit Brussel

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

This talk . . .

- given data is "noisy" (errors-in-variables setup)

  $\mathscr{I}_{\rm g}$ — given/specified elements of w

  $w|_{\mathscr{I}_{\rm g}}$ — selects the elements $\mathscr{I}_{\rm g}$ of w

- problem is to simultaneously
  - approximate $w|_{\mathscr{I}_{\rm g}}$ and
  - fill in the missing values

  by an LTI system of bounded complexity

- special case: exact identification with missing data

Exact identification with missing data

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

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

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

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

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

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

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

- and the problem is

$$\begin{array}{lll} \text{minimize} & \text{over } \widehat{w} & \| w - \widehat{w} \|_v \\ \text{subject to} & \text{rank}\left(\mathscr{H}_L(\widehat{w})\right) \leq r \end{array} \qquad \text{(SLRA)}$$

- weighted Hankel structured low-rank approximation

Parameter optimization

- using the kernel parameterization

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

$$\begin{array}{lcl} q & — & \text{\# of variables} \\ \mathtt{p} := qL - r & — & \text{co-rank (rank deficiency)} \end{array}$$

- (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

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

By assumption A3 the solution is unique.

About assumptions A1–A3

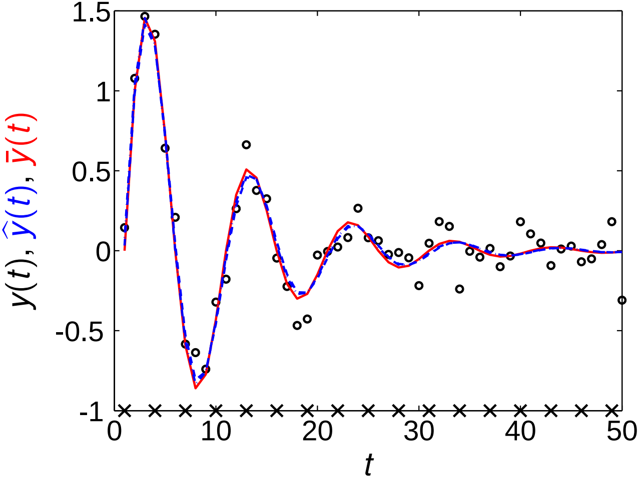
- A1 and A3 ensure uniqueness of y
- otherwise, $y \in B^+(c - Ax) + \text{null}(B)$
- A2 ensures feasibility with a nontrivial solution
- with $m = n_y$, (GLN) has trivial solution $f = 0$
- with $m - n_y > n_x$, (GLN) generically has no solution

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)

Autonomous system identification



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

Model-free simulation

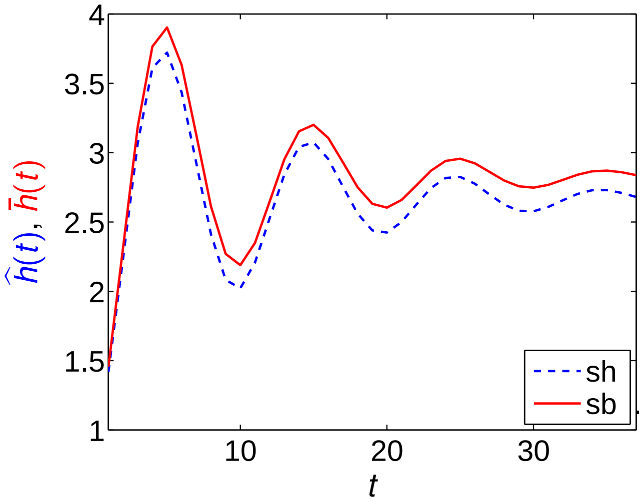
- second order SISO system, defined by

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

- $w^{(1)}$ is noisy trajectory generated from random input
- $w^{(2)}$ is the impulse response estimate $\bar{h}$, *i.e.*,

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

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



true impulse response $\bar{h}$

model free estimate $\widehat{h}$

Conclusions

- element-wise weighted approximation criterion
- zero weights $\leftrightarrow$ missing values
- cost function evaluation

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

- unique solution under certain assumptions
- nonlinear least-squares problem
- initial approximation from nuclear norm relaxaion