Exact 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

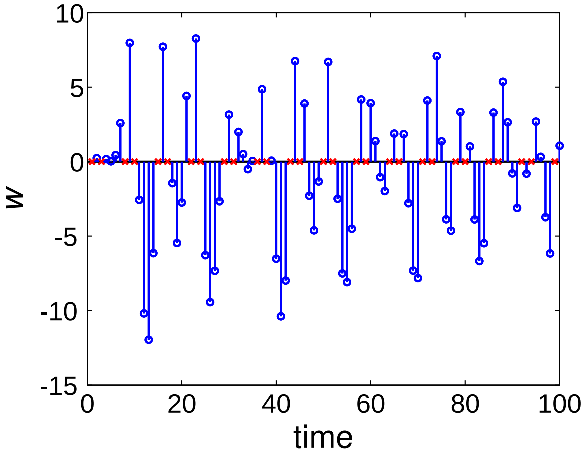
This talk . . .

- given data is 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 $h(1), \dots, h(T)$
  - missing data — extension $h(T+1), \dots$

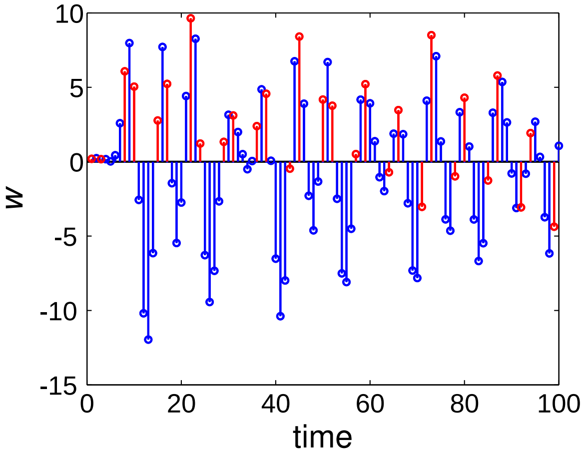
Example



○ — 6th order autonomous LTI system's trajectory

× — missing data locations

Example



○ — 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

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

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

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

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

Types of methods

- convex relaxations (nuclear norm heuristic)

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

  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 $\mathscr{H}_4(w)$

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

- and select the columns in blue and red

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

- 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

- 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

On the choice of L