Hankel structured low rank matrix completion

Ivan Markovsky

University of Southampton

Hankel matrices

an $m \times n$ matrix H is Hankel structured

$$\Updownarrow$$

there is a sequence $h = \big(h(1), \ldots, h(T)\big)$, $T := m + n - 1$

such that

$$H = \underbrace{\begin{bmatrix} h(1) & h(2) & \color{red}{h(3)} & \cdots & h(n) \\ h(2) & \color{red}{h(3)} & \iddots & \iddots & h(n+1) \\ \color{red}{h(3)} & \iddots & \iddots & & h(n+2) \\ \vdots & \iddots & & & \vdots \\ h(m) & h(m+1) & \cdots & \cdots & h(T) \end{bmatrix}}_{\color{red}{\mathscr{H}_m(h)}}$$

- the sequence h parameterizes the Hankel matrix H
- there is a bijection $(h,m) \leftrightarrow H$

$$(h,m) \xrightarrow{\mathscr{H}_m(h)} H \qquad \text{Hankel matrix constructor}$$

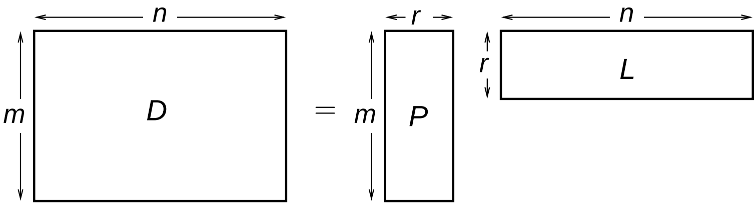
$$H \xrightarrow{\text{par}(H)} (h,m) \qquad \text{parameters readout map}$$

Low rank matrices

an $m \times n$ matrix D has rank less than or equal to r

$$\Updownarrow$$

there is an $m \times r$ matrix P and an $r \times n$ matrix L, such that



$$\Updownarrow$$

there is an $(m - r) \times m$ matrix R, such that

$$RR^\top = I_{m-r} \qquad \text{and} \qquad RD = 0$$

Low rank Hankel matrices

$$\operatorname{rank}\big(\mathscr{H}_L(h)\big) \leq r, \text{ for some } h = \big(h(1),\dots,h(T)\big)$$
and L, such that $L > r$ and $T - L + 1 \geq r$

$$\Updownarrow$$

there is $p_0, p_1, \dots, p_{r+1}$, such that

$$\sum_{\tau=0}^{r} p_\tau h(t+\tau) = 0, \quad \text{for } t = 1,\dots,T-L+1$$

Linear time-invariant autonomous systems

$$\mathscr{B}(p) = \left\{ h = \big(h(1), h(2) \ldots\big) \;\Big|\; \sum_{\tau=0}^{r} p_\tau h(t+\tau) = 0, \text{ for } t = 1, 2, \ldots \right\}$$

$$\Updownarrow$$

$$\mathscr{B}(A, c) = \left\{ h \;\Big|\; \exists\, b \in \mathbb{R}^r, \text{ such that } h(t) = cA^{t-1}b, \text{ for } t = 1, 2, \ldots \right\}$$

$h \in \mathscr{B}$	—	h is a trajectory of the model $\mathscr{B}$
$\dim(\mathscr{B}) = r$	—	model complexity
$\mathscr{L}$	—	the set of autonomous LTI models
$\mathscr{L}_r$	—	set of LTI models with complexity at most r

Low rank Hankel matrices

$\updownarrow$

LTI system trajectories

rank $\big(\mathscr{H}_m(h)\big) \leq r$, for some $h = \big(h(1),\dots,h(T)\big)$
and L, such that $L > r$ and $T - L + 1 \geq r$

$$\Updownarrow$$

there is $\mathscr{B} \in \mathscr{L}_r$, such that
$\big(h(1),\dots,h(T-L+1)\big)$ is a trajectory of $\mathscr{B}$

Most powerful unfalsified model

$\mathscr{B}$ is the most powerful unfalsified model of h

(in the model class $\mathscr{L}$)

$$\overset{..}{\Updownarrow}$$

1. $h \in \mathscr{B} \in \mathscr{L}$ (unfalsified)
2. $\dim(\mathscr{B})$ is minimal (most powerful)

$$\Updownarrow$$

minimal partial realization of h

notation: $\mathscr{B}_{\text{mpum}}(h)$

The identification problem

$$\text{data } h \xrightarrow{\text{identification method}} \text{model } \mathscr{B}$$

most powerful unfalsified model:

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{L} \quad \dim(\mathscr{B}) \quad \text{subject to} \quad h \in \mathscr{B}$$

$$\Updownarrow$$

$$\text{minimize} \quad \text{over } h(T+1),\dots,h(2T-1) \quad \text{rank}\left(\mathscr{H}_T(h_{\text{ext}})\right)$$

where

$$h_{\text{ext}} = \big(\underbrace{h(1),\dots,h(T)}_{h},\underbrace{h(T+1),\dots,h(2T-1)}_{\text{optimization variables}}\big)$$

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{L} \quad \dim(\mathscr{B}) \quad \text{subject to} \quad h \in \mathscr{B}$$

$$\Updownarrow$$

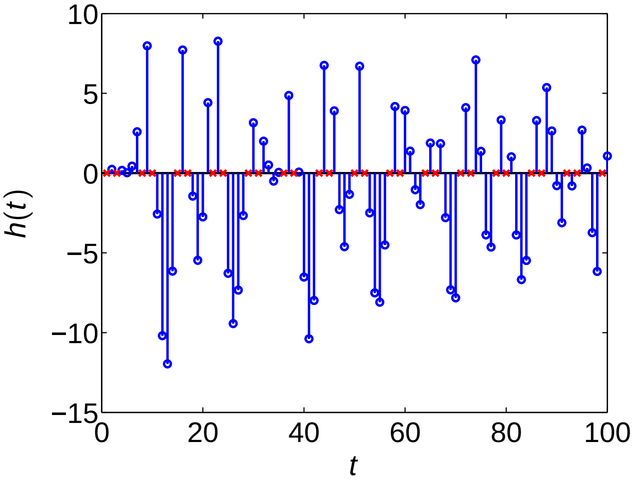
$$\text{minimize} \quad \text{over the ?'s} \quad \operatorname{rank} \begin{bmatrix} h(1) & h(2) & h(3) & \cdots & h(T) \\ h(2) & h(3) & \iddots & h(T) & ? \\ h(3) & \iddots & \iddots & \iddots & ? \\ \vdots & h(T) & \iddots & \iddots & \vdots \\ h(T) & ? & ? & \cdots & ? \end{bmatrix}$$

Missing data

- partial realization
  - $\rightsquigarrow$ missing (to be estimated) values in the "future"
  - $\rightsquigarrow$ extrapolation of sequence by an aut. LTI system
- generalization:

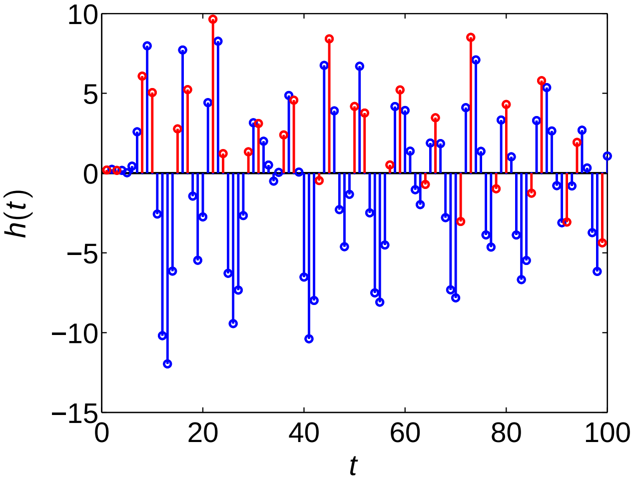
  find $\mathscr{B}_{\text{mpum}}(h)$ when arbitrary values of h are missing
  - $\rightsquigarrow$ interpolation/extrapolation of a sequence by (multivariable, input/output) LTI model
  - $\rightsquigarrow$ Hankel structured low rank matrix completion problem

Example



data points missing values locations

Example



data points interpolated values

Missing data problem

$\mathscr{I}_{\text{data}}$ — set of indexes of the given values of h
L — upper bound on the model complexity

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{h} \quad \operatorname{rank}\big(\mathscr{H}_L(\widehat{h})\big) \\ \text{subject to} & \widehat{h}(\mathscr{I}_{\text{data}}) = h(\mathscr{I}_{\text{data}}) \end{array}$$

solution methods:

- nuclear norm heuristic
- subspace methods
- local optimization methods

Nuclear norm heuristic for matrix completion

$\|A\|_* :=$ sum of the singular values of A

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

$$\Updownarrow$$

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

semidefinite programming problem

CVX code

```
function hh = hmc(h, L)

T   = length(h);
I_m = find( isnan(h));
I_d = find(~isnan(h));

cvx_begin sdp;
  variable U(L, L)                         symmetric;
  variable V(T - L + 1, T - L + 1) symmetric;
  variable h_m(length(I_m), 1);
  minimize(trace(U) + trace(V));
  subject to
    hh(I_m) = h_m; hh(I_d) = h(I_d);
    H = hankel(hh(1:L), hh(L:end));
    [U H'; H V] > 0;
cvx_end
```

Numerical example

```
n = 3; T = 10; sys0 = drss(n);

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

hh = hmc(h, T)'; norm(h0 - hh)
```

```
Calling SDPT3: 210 variables, 91 equality constraints
------------------------------------------------------------

 num. of constraints = 91
 dim. of sdp    var  = 20,   num. of sdp  blk  =  1
*******************************************************************
   SDPT3: Infeasible path-following algorithms
*******************************************************************
 version  predcorr  gam  expon  scale_data
   HKM      1      0.000   1        0
it pstep dstep pinfeas dinfeas  gap      mean(obj)   cputime
-------------------------------------------------------------------
 0|0.000|0.000|1.6e-01|3.7e+00|1.5e+03| 1.337736e+02| 0:0:00| chol  1  1
 1|1.000|1.000|2.2e-08|8.2e-02|1.9e+02| 8.595944e+01| 0:0:00| chol  1  1
 2|0.979|1.000|3.1e-08|8.2e-03|4.8e+00| 2.415990e+00| 0:0:00| chol  1  1
 3|1.000|0.953|9.4e-09|1.2e-03|1.5e+00| 1.145234e+00| 0:0:00| chol  1  1
 4|0.976|1.000|1.5e-09|8.2e-05|6.0e-02| 5.225036e-01| 0:0:00| chol  1  1
 5|1.000|1.000|4.0e-10|8.2e-06|9.4e-03| 5.022662e-01| 0:0:00| chol  1  1
 6|0.970|0.978|1.2e-10|9.8e-07|2.9e-04| 4.991459e-01| 0:0:00| chol  1  1
 7|0.985|0.987|1.3e-10|9.3e-08|4.2e-06| 4.990335e-01| 0:0:00| chol  1  1
 8|0.999|1.000|5.3e-10|2.6e-11|7.9e-08| 4.990318e-01| 0:0:00| chol  1  1
 9|1.000|1.000|2.2e-09|4.0e-11|4.6e-09| 4.990318e-01| 0:0:00|
  stop: max(relative gap, infeasibilities) < 1.49e-08
-------------------------------------------------------------------
 number of iterations   =  9
 ...... more output ......
 Total CPU time (secs)  = 0.3
 CPU time per iteration = 0.0
 termination code       =  0
-------------------------------------------------------------------
Status: Solved
Optimal value (cvx_optval): +0.499032

ans =   5.6272e-11
```

Subspace method for identification with missing data

- let $\operatorname{rank}\big(\mathscr{H}_{r+1}(h)\big) = r$
- there is $p \neq 0$, such that

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_r \end{bmatrix} \mathscr{H}_{r+1}(h) = 0$$

- consider the matrix $\mathscr{H}_{r+2}(h)$; we have

$$\underbrace{\begin{bmatrix} p_0 & p_1 & \cdots & p_r & 0 \\ 0 & p_0 & p_1 & \cdots & p_r \end{bmatrix}}_{\widetilde{P}} \mathscr{H}_{r+2}(h) = 0$$

- $\widetilde{P}$ is full row rank, so that for any i there is $\widetilde{p}^i \neq 0$, such that

$$\widetilde{p}^i \mathscr{H}_{r+2}(h) = 0 \qquad \text{and} \qquad \widetilde{p}^i_i = 0$$

- suppose that $\mathscr{H}_{r+2}(h)$ has at least $r+1$ columns with single missing value in the ith position

- denote the corresponding submatrix of $\mathscr{H}_{r+2}(h)$ by $\widetilde{H}^i$

- left $\ker(\widetilde{H}^i) = \alpha \widetilde{p}^i$, for some $\alpha \neq 0$

  $\rightsquigarrow$ $\widetilde{p}^i$ can be identified (up to scaling factor) from h

- suppose also that $\mathscr{H}_{r+2}(h)$ has at least $r+1$ columns with single missing value in the jth position, where $j \neq i$

  $\rightsquigarrow$ $\widetilde{p}^j$ can be identified (up to scaling factor) from h

- $\mathrm{GCD}(p^i, p^j) = p$ $\rightsquigarrow$ identification algorithm with missing values

Extensions

- $\Delta > 1$ missing value per frame $\rightsquigarrow$ $\mathscr{H}_{r+\Delta}(h)$
- multivariable time series $h(t) \in \mathbb{R}^{\mathrm{p}}$

  $\rightsquigarrow$ block-Hankel matrix of rank $=$ n (order of the system)
- open systems $w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$, u — input, y — output

  $\rightsquigarrow$ block-Hankel matrix of rank $=$ rank $\big(\mathscr{H}_L(u)\big) +$ n
- multiple time series

$$w = \{w^1, \dots, w^N\}, \quad \text{where} \quad w^i = \big(w^i(1), \dots, w^i(T^i)\big)$$

  $\rightsquigarrow$ block partitioned, block-Hankel matrix

$$\begin{bmatrix} \mathscr{H}_L(w^1) & \cdots & \mathscr{H}_L(w^N) \end{bmatrix}$$

Open systems

$$\text{rank}\left(\mathscr{H}_L(w)\right) \leq \text{rank}\left(\mathscr{H}_L(u)\right) + \mathtt{n}, \text{ for some}$$

$$w = \big(w(1), w(2), \ldots, w(T)\big), \quad w = \begin{bmatrix} u \\ y \end{bmatrix}, \quad \dim(u) = \mathtt{m}$$

$$\text{and } L, \text{ such that } L > r \text{ and } T - L + 1 \geq r$$

$$\Updownarrow$$

$$\big(w(1), \ldots, w(T - L + 1)\big) \in \mathscr{B} \in \mathscr{L}_{\mathtt{m},\mathtt{n}}$$

$\mathscr{L}_{\mathtt{m},\mathtt{n}}$ — class of LTI systems with $\leq \mathtt{m}$ inputs and order $\leq \mathtt{n}$

Multiple time series

$$\text{rank}\left(\begin{bmatrix}\mathscr{H}_L(w^1) & \cdots & \mathscr{H}_L(w^N)\end{bmatrix}\right) \leq \text{rank}\left(\mathscr{H}_L(u)\right) + \text{n}, \text{ for some}$$

$$w^i = \left(w^i(1), w^i(2), \ldots, w^i(T^i)\right), \quad w^i = \begin{bmatrix} u^i \\ y^i \end{bmatrix}, \quad \dim(u^i) = \mathtt{m}$$

and L, such that $L > r$ and $T - L + 1 \geq r$

$$\Updownarrow$$

$$\left(w^i(1), \ldots, w^i(T^i - L + 1)\right) \in \mathscr{B} \in \mathscr{L}_{\mathtt{m},\mathtt{n}}, \quad \text{for } i = 1, \ldots, N$$

Applications

- partial realization
- data-driven simulation
- data-driven tracking
- state estimation (Kalman filter) with missing data

Classical simulation problem

given

- system $\mathscr{B} \in \mathscr{L}$ (specified by some representation)
- initial condition w_{ini}, and (specified by a trajectory of $\mathscr{B}$)
- input u,

find the output y of $\mathscr{B}$, corresponding to w_{ini} and u, *i.e.*,

$$w_{\text{ini}} \wedge (u, y) \in \mathscr{B}$$

- there are many ways to solve the problem
- the algorithms depend on the model representation (state-space, transfer function, impulse response, ...)

Data-driven simulation

given

- trajectory w_{d} of a system $\mathscr{B} \in \mathscr{L}$
- initial condition w_{ini}, and
- input u,

find the output y of $\mathscr{B}$, corresponding to w_{ini} and u, *i.e.*,

$$w_{\mathrm{ini}} \wedge (u, y) \in \mathscr{B}$$

$$\Updownarrow$$

matrix completion problem for block partitioned Hankel matrix

$$\text{minimize} \quad \text{over } y \quad \operatorname{rank}\left(\begin{bmatrix} \mathscr{H}_L(w_{\mathrm{d}}) & \mathscr{H}_L\big((u, y)\big) \end{bmatrix}\right)$$

Classical tracking problem

given

- system $\mathscr{B} \in \mathscr{L}$ (specified by some representation)
- initial condition w_{ini}, and (specified by a trajectory of $\mathscr{B}$)
- desired output y,

find an input u of $\mathscr{B}$, achieving perfect tracking, *i.e.*,

$$w_{\text{ini}} \wedge (u, y) \in \mathscr{B}$$

- again there are different ways to solve the problem
- the algorithms depend on the model representation (state-space, transfer function, impulse response, ...)

Data-driven tracking

given

- trajectory w_d of a system $\mathscr{B} \in \mathscr{L}$
- initial condition w_ini, and
- desired output y,

find an input u of $\mathscr{B}$, achieving perfect tracking, *i.e.*,

$$w_\text{ini} \wedge (u, y) \in \mathscr{B}$$

$$\Updownarrow$$

matrix completion problem for block partitioned Hankel matrix

$$\text{minimize} \quad \text{over } u \quad \text{rank}\left(\begin{bmatrix} \mathscr{H}_L(w_\text{d}) & \mathscr{H}_L((u, y)) \end{bmatrix}\right)$$

Interpolation with given model

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_r \end{bmatrix} \mathscr{H}_{r+1}(w) = 0$$

$$\Updownarrow$$

$$\underbrace{\begin{bmatrix} R_0 & R_1 & \cdots & R_r & & & \\ & R_0 & R_1 & \cdots & R_r & & \\ & & \ddots & \ddots & & \ddots & \\ & & & R_0 & R_1 & \cdots & R_r \end{bmatrix}}_{A} \begin{bmatrix} w(1) \\ w(2) \\ \vdots \\ w(T) \end{bmatrix} = 0$$

$$\Updownarrow$$

$$A(:, \mathscr{I}_{\text{missing}}) w(\mathscr{I}_{\text{missing}}) = A(:, \mathscr{I}_{\text{data}}) w_{\text{data}}$$

note: recursive solution $\rightsquigarrow$ growing set of unknowns
(initial conditions and missing inputs)

Conclusions

- LTI system trajectories $\leftrightarrow$ low rank Hankel matrices
- LTI model identification from trajectory with missing values $\leftrightarrow$ low rank Hankel structured matrix completion problem
- nuclear norm heuristic and subspace type methods
- the subspace method is applicable when part of the Hankel matrix completely specifies the model
- the methods generalize to multivariable open systems and multiple time series with missing values
- applications for partial realization, data-driven simulation and control

Inexact data

- inexact data $\rightsquigarrow$ approximation of incomplete matrix

$$\text{minimize} \quad \text{over } \widehat{h} \quad \begin{bmatrix} \operatorname{rank}\left(\mathscr{H}_L(\widehat{h})\right) \\ \|\widehat{h}(\mathscr{I}_{\text{data}}) - h(\mathscr{I}_{\text{data}})\| \end{bmatrix} \quad \begin{matrix} \leftarrow \text{ low complexity} \\ \leftarrow \text{ good fit} \end{matrix}$$

- scalarizations of the biobjective problem

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{h} \quad \|\widehat{h}(\mathscr{I}_{\text{data}}) - h(\mathscr{I}_{\text{data}})\| \\ &\text{subject to} \quad \operatorname{rank}\left(\mathscr{H}_{r+1}(\widehat{h})\right) \leq r \quad \leftarrow \text{ complexity bound} \end{aligned}$$

or

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{h} \quad \operatorname{rank}\left(\mathscr{H}_L(\widehat{h})\right) \\ &\text{subject to} \quad \|\widehat{h}(\mathscr{I}_{\text{data}}) - h(\mathscr{I}_{\text{data}})\| \leq e \quad \leftarrow \text{ error bound} \end{aligned}$$

Inexact data

- another scalarization of the biobjective problem

$$\text{minimize} \quad \text{over } \widehat{h} \quad \operatorname{rank}\big(\mathscr{H}_L(\widehat{h})\big) + \gamma \|\widehat{h}(\mathscr{I}_{\text{data}}) - h(\mathscr{I}_{\text{data}})\|$$

↑ trade-off parameter

- all problems “sweep” the set of Pareto optimal solutions
- replacing rank by nuclear norm $\rightsquigarrow$ convex relaxations
- can be solved by local optimization methods

Questions?