Regularized structured low-rank approximation

Vrije Universiteit Brussel

Mariya Ishteva Konstantin Usevich Ivan Markovsky Dept. ELEC, VUB

erc

Hankel structure in system identification

- Given response $y = [y(1), \dots, y(T)] \in \mathbb{R}^T$ of
  an autonomous dynamic discrete-time LTI system, order n.
  Estimate model parameters for some representation of the system.

- Difference equation:
  $\theta_0\, y(t) + \theta_1\, y(t+1) + \cdots + \theta_n\, y(t+n) = 0, \quad \text{for } t = 1, \dots, T-n.$

$$[\theta_0\ \theta_1 \cdots \theta_n] \begin{bmatrix} y(1) & y(2) & y(3) & \dots\, y(T-n) \\ y(2) & y(3) & y(4) & \iddots \\ y(3) & \iddots & & \vdots \\ \vdots & \iddots & \iddots & \\ y(n+1) & y(n+2) & y(n+3) \dots & y(T) \end{bmatrix} = 0$$

- Equivalently, the matrix is rank deficient!
  Noisy data: Hankel low-rank approximation of a Hankel matrix.

Approximate greatest common divisor

- Given 3 polynomials
  $$a(z) = a_0 + a_1 z + \cdots + a_{n_1} z^{n_1},$$
  $$b(z) = b_0 + b_1 z + \cdots + b_{n_2} z^{n_2},$$
  $$c(z) = c_0 + c_1 z + \cdots + c_{n_3} z^{n_3}.$$
  Find their greatest common divisor.

- $a(z)$, $b(z)$, and $c(z)$ have a common divisor if

$$\mathit{rank} \begin{bmatrix} S_{n_1}(b) & S_{n_1}(c) \\ S_{n_2}(a) & \mathbf{0} \\ \mathbf{0} & S_{n_3}(a) \end{bmatrix} < n_1 + n_2 + n_3. \qquad \left.\begin{bmatrix} a_0 & a_1 & \cdots & a_{n_1} & & \mathbf{0} \\ & \ddots & \ddots & & \ddots & \\ \mathbf{0} & & a_0 & a_1 & \cdots & a_{n_1} \end{bmatrix}\right\} n_3$$

- Equivalently, the matrix is rank deficient!
  Inexact coefficients: structured low-rank approximation.

Affinely structured matrices

(Block) Hankel/Toeplitz matrices,
Sylvester, fixed sparsity,
banded matrices,
unstructured matrices.

$$\mathcal{S}(p) = \begin{bmatrix} p_1 & p_4 & p_3 & p_6 & p_1 & 0 \\ p_2 & p_3 & 0 & p_3 & p_2 & p_6 \\ p_1 & 0 & p_7 & p_3 & 0 & p_3 \\ p_3 & p_5 & p_6 & 0 & p_7 & p_2 \end{bmatrix}$$

Structure preserving low-rank approximation

$$\min_{\hat p} \|p - \hat p\|_2, \quad \text{subject to} \quad \operatorname{rank}(\mathcal{S}(\hat p)) \le r.$$

- Low-rank and structure are easily handled separately
  - SVD
  - Orthogonal projection of a given matrix on the set of structured matrices $\mathcal{S}$
    $\leftrightarrow$ Averaging elements corresponding to the same parameter.
    Notation: $P_{\mathcal{S}}(X) = \mathcal{S}(\underbrace{\Pi_{\mathbf{S}} \operatorname{vec}(X)}_{p_X})$.
- However, imposing both is nontrivial.
- Existing approaches: kernel representation: $\exists R : R\mathcal{S}(\hat p) = 0$.

Image representation

- Given: a structured matrix D.
- Find: two factors P and L (with reduced dimension) , s.t.

$D \approx PL$ and PL is a structured matrix.

- Challenge: how to impose the structure via the factors?

Enforcing structure by regularization

$\to$ low-rank
$\to$ structure

$\to$ low-rank
$\to$ *regularized* structure

$\to$ *regularized* low-rank
$\to$ structure

$\min_{P,L} \|D - PL\|_W^2 + \lambda \|PL - P_{\mathcal{S}}(PL)\|_F^2$

weight matrix   regularization parameter

$\min_{P,L} \|D - P_{\mathcal{S}}(PL)\|_W^2 + \lambda \|PL - P_{\mathcal{S}}(PL)\|_F^2$

$\equiv$

$$\min_{P,L} \ \|p - \underbrace{\Pi_{\mathbf{S}} \operatorname{vec}(PL)}_{p_{PL}}\|_2^2 + \lambda \|PL - P_{\mathcal{S}}(PL)\|_F^2$$

- For $\lambda = \infty$ the three problems are equivalent.

Algorithm

Iterate (until convergence)

- Iterate (until convergence)
  - Improve the approximation of L:

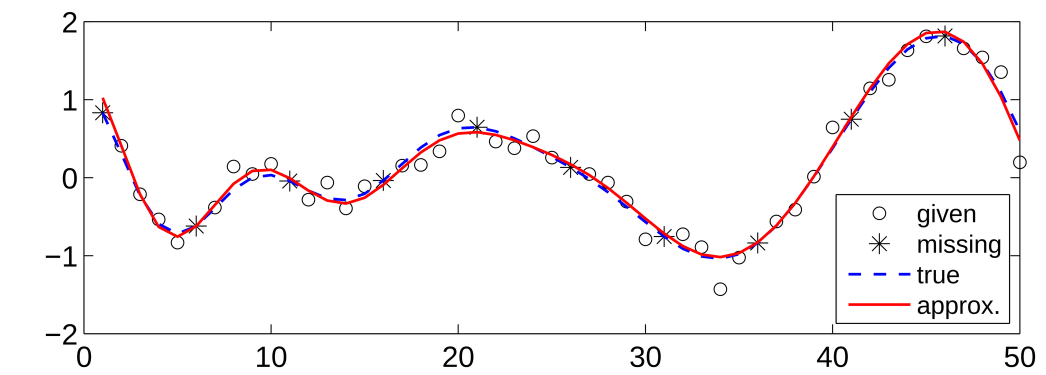
$$L = \arg\min_L \left\| \begin{bmatrix} \Pi_{\mathbf{S}} \\ \sqrt{\lambda}(I_{mn} - \mathbf{S}\,\Pi_{\mathbf{S}}) \end{bmatrix} (I_n \otimes P) \operatorname{vec}(L) - \begin{bmatrix} p \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

  $\to$ Least squares problem ($Ax \approx b$); easily solved by standard techniques.
  - Improve the approximation of P
- Increase λ.

Examples

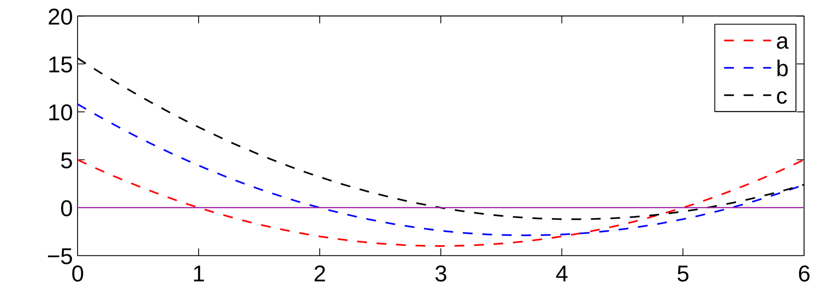
System identification with missing data

- True rank: 4 $\to$ need 5 consecutive data points for identification
- However below, every 5th data point is missing

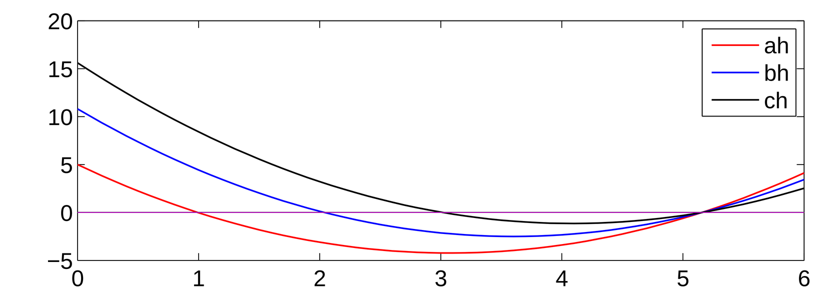


Approximate greatest common divisor

$a(z) = (1 - z)(5 - z),$
$b(z) = (2 - z)(5.4 - z),$
$c(z) = (3 - z)(5.2 - z).$



$\hat a(z) = (0.99 - z)(5.14 - z),$
$\hat b(z) = (2.04 - z)(5.14 - z),$
$\hat c(z) = (3.02 - z)(5.14 - z).$



Discussion

+ general and simple approach
+ addresses a missing point of view: image representation
+ weighted problems: $\min_{\hat p} \|p - \hat p\|_w^2$
+ fixed values, missing values

− how to optimally increase λ?

−/+ computational cost / lower cost for smaller targeted rank

→ other constraints (e.g., smoothness)