Regularized structured low-rank approximation

Mariya Ishteva   Konstantin Usevich   Ivan Markovsky

Dept. ELEC, Vrije Universiteit Brussel

Benelux Meeting on Systems and Control 2013

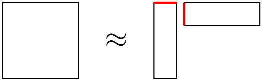
Outline

Introduction: low rank, structure

Regularized structured low-rank approximation

Example and discussion

Low-rank approximation



- Problem: approximate a matrix by a low-rank matrix
- Used in: data mining, machine learning, signal processing
- Used for: dimensionality reduction and factor analysis
- Generalization: structured low-rank approximation

Structured matrices

- relation between elements: $\mathscr{S} \longleftrightarrow p$
  - pattern (linear structure)
  - nonlinear structure

$$\begin{bmatrix} p_1 & p_2 & p_3 & \cdots & p_n \\ p_2 & p_3 & p_4 & \iddots & \\ p_3 & \iddots & & & \vdots \\ \vdots & \iddots & \iddots & & \\ p_m & p_{m+1} & p_{m+2} & \cdots & p_{n_p} \end{bmatrix}, \qquad \begin{bmatrix} p_1 & p_4 & p_3 & p_6 & 0 \\ p_2 & p_3 & 0 & p_3 & p_2 \\ p_1 & 0 & p_7 & p_3 & 0 \\ p_3 & p_5 & p_6 & 0 & p_7 \\ 0 & p_6 & p_1 & p_4 & 0 \end{bmatrix}$$

- (block) Hankel/Toeplitz, Sylvester, banded matrix
- sparse matrices (with fixed sparsity pattern)
- unstructured matrices

Hankel structure in system identification

Example: autonomous dynamic discrete-time linear time-invariant system, order n, single output

- Given response $y = [y(1), \ldots, y(T)] \in \mathbb{R}^T$ of a system
- Choose a representation of the system;

  Estimate model parameters

Difference equation:

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

Hankel structure in system identification

Difference equation:

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

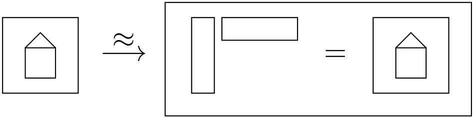
$$\begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix} \begin{bmatrix} y(1) & y(2) & y(3) & \ldots & y(T-n) \\ y(2) & y(3) & y(4) & \iddots & \\ y(3) & \iddots & & & \vdots \\ \vdots & \iddots & \iddots & & \\ y(n+1) & y(n+2) & y(n+3) & \ldots & y(T) \end{bmatrix} = 0$$

Equivalently, the matrix is rank deficient!

Noisy data: Hankel low-rank approximation of a Hankel matrix.

Structured low-rank approximation

Problem: structure preserving low-rank approximation



Outline

Introduction: low rank, structure

Regularized structured low-rank approximation

Example and discussion

Optimization problem

Structured low-rank approximation:

$$\min_{\hat{p}} \|p - \hat{p}\|_2 , \quad \text{such that} \quad \operatorname{rank}(\mathscr{S}(\hat{p})) \leq r .$$

Low-rank and structure can easily be handled separately

- SVD
- orthogonal projection

However, imposing both is nontrivial.

Orthogonal projection

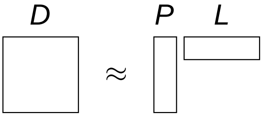
- For Hankel matrices:
  averaging elements corresponding to the same parameter.

- True for a whole class of affine structures,
  - including (block) Hankel/Toeplitz, Sylvester matrices,
  - each matrix element corresponds to only one element of the parameter vector.

- Let $\mathrm{vec}(X)$ denote the vectorized matrix X.

  Notation: $P_{\mathscr{S}}(X) = \mathscr{S}(\underbrace{\Pi_{\mathbf{S}} \mathrm{vec}(X)}_{p_X})$.

Image representation

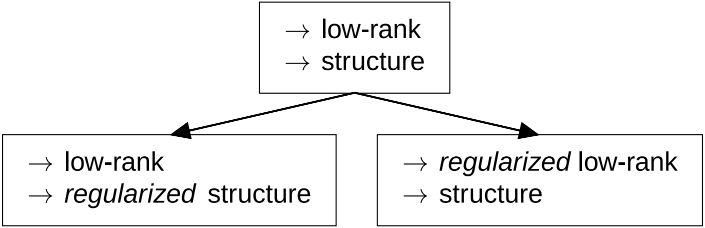


- Given: a structured matrix $D \in \mathbb{R}^{m \times n}$ and r, s.t. $r \ll m, n$,
- Find: two factors $P \in \mathbb{R}^{m \times r}$ and $L \in \mathbb{R}^{r \times n}$, s.t.

$$D \approx PL \quad \text{and} \quad PL \text{ is a structured matrix.}$$

- Challenge: how to impose the structure via the factors?
  - for Hankel matrices: use nonlinear Vandermonde structure
  - for general linear structures: ?

Enforcing structure by regularization

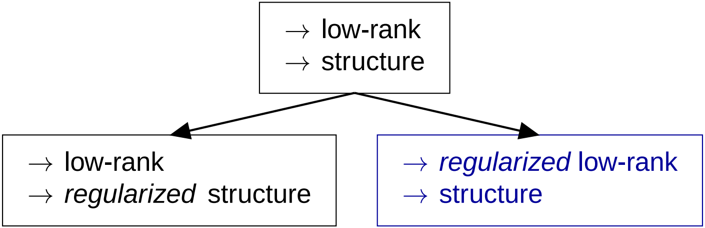


$$\min_{P,L} \|D - PL\|_W^2 + \lambda \|PL - P_{\mathscr{S}}(PL)\|_F^2 \qquad \min_{P,L} \|D - P_{\mathscr{S}}(PL)\|_W^2 + \lambda \|PL - P_{\mathscr{S}}(PL)\|_F^2$$

weight matrix — regularization parameter — Frobenius norm

- For $\lambda = \infty$ the three problems are equivalent.
- Different interpretations
- In both cases both constraints are satisfied at the solution.

Optimization problem



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

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

High-level idea

Alternatingly improve the approximations of P and of L

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

$$\min_{P} \|p - \Pi_{\mathbf{S}} \operatorname{vec}(PL)\|_2^2 + \lambda \|PL - P_{\mathscr{S}}(PL)\|_F^2.$$

Details

Let I_n be the $n \times n$ identity matrix and '$\otimes$' denote the Kronecker product. Using $\text{vec}(XYZ) = (Z^\top \otimes X)\text{vec}(Y)$ we have

$$\text{vec}(PL) = (I_n \otimes P)\,\text{vec}(L).$$

Now, $\min_L \|p - \Pi_{\mathbf{S}} \text{vec}(PL)\|_2^2 + \lambda \|PL - P_{\mathscr{S}}(PL)\|_F^2$, becomes

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

$$Ax \approx b.$$

- Least squares problem; easily solved by standard techniques
- Similarly for P

Algorithm

Input: Parameter vector $p \in \mathbb{R}^{n_p}$, rank $r \in \mathbb{N}$, $P_0 \in \mathbb{R}^{m \times r}$.
Output: Factors $P \in \mathbb{R}^{m \times r}$ and $L \in \mathbb{R}^{r \times n}$.

```
for j = 1, 2, ... do
    for k = 1, 2, ... do
        Compute L:
```

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

```
        Compute P:
```

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

```
    end for
    Set λ_{j+1} such that λ_{j+1} > λ_j.
end for
```

Details / issues

- For faster convergence:
  - start with small λ
  - increase with each step
- Initial guess P_0 - can be obtained from SVD
- We declare that PL is a structured matrix if

$$\|PL - P_{\mathscr{S}}(PL)\|_F^2 < 10^{-12}.$$

- High computational cost;
  can be reduced if we take advantage of the structure

Generalizations

- weighted problems: $\min_{\hat{p}} \|p - \hat{p}\|_w^2$
- general affine structures
- fixed values in the structure/approximation
- missing values in the given matrix
- other constraints, e.g., smoothness

Outline

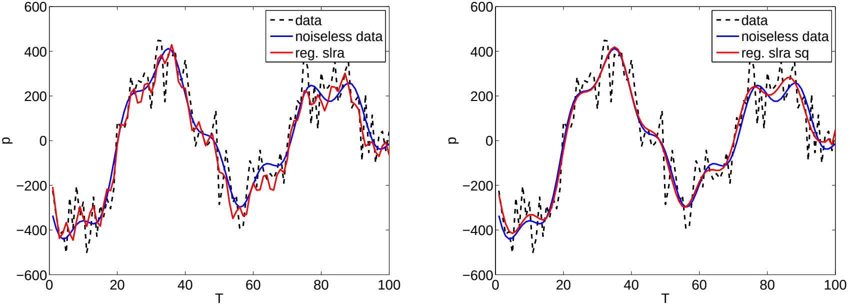
Introduction: low rank, structure

Regularized structured low-rank approximation

Example and discussion

Autonomous LTI dynamical system

Given: noisy data $p \in \mathbb{R}^{100}$, true $r = 6$



size (m,n):	$(7,94)$	$(50,51)$
speed:	faster	slower
$\|p-\hat{p}\|$:	854.1568 (smaller)	859.1257
$\|p_0-\hat{p}\|$:	512.4965	298.6408 (smaller)
$\|p-\hat{p}\|$ (5 runs):	958.2705	934.2849 (smaller)
$\|p_0-\hat{p}\|$ (5 runs):	512.0792	291.0884 (smaller)
shape:	less “smooth”	“smoother” (often)

Discussion

Advantages:

- addresses a missing point of view: image representation
- lower computational cost for smaller targeted rank
- any rank (existing approaches have some restrictions)

Disadvantages:

- how to optimally increase λ?

Open question:

- can we make use of the reshapable structures, e.g., for better initialization?

Summary

- A new algorithm for structured low-rank approximation
- Based on image representation; using regularization
- More efficient for small ranks
- Easily generalizable: affine structures, weights, fixed elements

Future work:

- Explore structure in solving the least squares problem
- Study convergence properties
- Generalize to handle missing data and extra constraints

Thank you!

Questions?

mariya.ishteva@vub.ac.be