ETH zürich

Automatic Control Laboratory

On Low-Rank Hankel Matrix Denoising

Mingzhou Yin, Prof. Roy S. Smith

July 14, 2021, SYSID 2021

Towards low-order structures

- One interpretation of system identification:
  Find a low-order description that is close to the collected data
- Low-order description $\rightarrow$ **rank deficiency** in data matrices

- **Example 1:** impulse response of discrete-time LTI systems

$$H_g = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-1} \\ g_1 & g_2 & \cdots & g_n \\ \vdots & \vdots & \ddots & \vdots \\ g_{m-1} & g_m & \cdots & g_{N-1} \end{bmatrix}$$ has a rank of the system order n_x.

- Applications in frequency-domain subspace identification (McKelvey 1996) and model order reduction (Markovsky 2005)

- **Example 2:** input-output trajectory of discrete-time LTI systems

$$U = \begin{bmatrix} u_0 & u_1 & \cdots & u_{n-1} \\ u_1 & u_2 & \cdots & u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_{m-1} & u_m & \cdots & u_{N-1} \end{bmatrix}, \quad Y = \begin{bmatrix} y_0 & y_1 & \cdots & y_{n-1} \\ y_1 & y_2 & \cdots & y_n \\ \vdots & \vdots & \ddots & \vdots \\ y_{m-1} & y_m & \cdots & y_{N-1} \end{bmatrix}$$

  If inputs are persistently excited, rank $\left(\begin{bmatrix} U \\ Y \end{bmatrix} \right) = mn_u + n_x$

- Applications in time-domain subspace identification (Moonen 1989) and data-driven simulation/control (Markovsky 2006)

The usual way…

- **Problem:** Estimate low-rank data matrix X from noisy measurement $W = X + \sigma Z$

- Find the closest low-rank **approximation** to the noisy data matrix

$$\hat{X}_{\text{LRA}} = \begin{array}{cc} \underset{\hat{X}}{\text{argmin}} & \left\| W - \hat{X} \right\|_F^2 \\ \text{s.t.} & \text{rank}(\hat{X}) \leq r. \end{array}$$

- Solution is given by the **Eckart-Young-Mirsky (EYM) theorem**

Truncated singular value decomposition

Let the singular value decomposition of W be $W = \sum_{i=1}^{m} w_i \mathbf{u}_i \mathbf{v}_i^\mathsf{T}$.

$$\hat{X}_{\mathsf{TSVD}} = \sum_{i=1}^{r} w_i \mathbf{u}_i \mathbf{v}_i^\mathsf{T}.$$

- Basic idea in principal component analysis / proper orthogonal decomposition
- When r is unknown, estimate r by inspection of scree plot or cross validation

Introducing Hankel structure

- Data matrices often have structural constraints $\rightarrow \hat{X}$ should also be structured
- **Generalized low-rank Hankel structure:** X is Hankel, rank$(X\Pi) = r$
- Covers Hankel matrix, Toeplitz matrix & Hankel matrices with noise-free rows
  - (Example 1) $X = H_g$, $\Pi = \mathrm{I}$, $r = n_x$
  - (Example 2, output noise only) $X = Y$, $\Pi = \mathrm{I} - U^\top(UU^\top)^{-1}U$ spans the null space of U, $r = n_x$
- **Structured** low-rank approximation (SLRA) problem

$$\hat{X}_{\text{SLRA}} = \underset{\hat{X}\in\mathcal{H}^{m\times n}}{\text{argmin}} \quad \left\|W - \hat{X}\right\|_F^2 \quad \text{s.t.} \quad \text{rank}(\hat{X}\Pi) \leq r.$$

- EYM theorem no longer valid $\rightarrow$ no closed-form solution

Solving the SLRA problem

- Iterative structural approximation (Wang 2019, Li 1997)

Iterative algorithm for SLRA

1: $W_1 \leftarrow W$
2: **repeat**
3: $\quad W_2 \leftarrow \hat{X}_{\text{TSVD}}(W_1)$
4: $\quad W_1 \leftarrow \mathcal{H}(W_2)$
5: **until** $\|W_1 - W_2\| < \epsilon \|W_1\|$
6: **Output:** $\hat{X} = W_1$

$\mathcal{H}(\cdot)$: orthogonal projector onto Hankel matrix set by averaging skew diagonals

- Nonlinear local optimization (Markovsky 2013)
- Relaxation by nuclear norm regularization (Fazel 2001)

$$\hat{X}_{\text{nuc}} = \underset{\hat{X} \in \mathcal{H}^{m \times n}}{\text{argmin}} \frac{1}{2} \left\| W - \hat{X} \right\|_F^2 + \tau \left\| \hat{X} \Pi \right\|_*$$

Approximation is NOT denoising

- The true objective is to minimize

$$\text{MSE}(\hat{X}) := \mathbb{E}\left(\left\|X - \hat{X}\right\|_F^2\right)$$

  instead of $\left\|W - \hat{X}\right\|_F^2$

- **An extreme case:** When $\sigma \to \infty$, $\hat{X}_{\text{TSVD}} \to \infty$ while min-MSE solution is zero
- **Problem:** Noise matrix also inflates non-zero singular values

$$\lim_{n\to\infty} w_i = \begin{cases} D_{\mu_Z}^{-1}(1/x_i^2), & x_i^2 > 1/D_{\mu_Z}(b^+) \\ b, & x_i^2 \le 1/D_{\mu_Z}(b^+) \end{cases}$$

- $w_i > x_i, \forall x_i$; no hope to recover modes with small s.v.

Singular value shrinkage

$$\hat{X}_{\text{shrink}} = \sum_{i=1}^{m} \eta(w_i)\mathbf{u}_i\mathbf{v}_i^{\mathsf{T}} + W(\mathsf{I}_n - \Pi), \;\; \eta(w_i) \in [0, w_i]$$

- Shrinkage law with minimum asymptotic MSE (Nadakuditi 2014)

$$\eta(w;\mu_Z) = \begin{cases} -2\frac{D_{\mu_Z}(w)}{D'_{\mu_Z}(w)}, & D_{\mu_Z}(w) < D_{\mu_Z}(b^+) \\ 0, & D_{\mu_Z}(w) \geq D_{\mu_Z}(b^+) \end{cases},$$

- When Z has i.i.d Gaussian entries, μ_Z has analytical solution (Marchenko-Pastur distribution)
- Optimal shrinkage law (Gavish 2014)

$$\eta(w) = \begin{cases} \frac{n\sigma^2}{w}\sqrt{\left(\frac{w^2}{n\sigma^2} - \beta - 1\right)^2 - 4\beta}, & w > (1+\sqrt{\beta})\sqrt{n}\sigma \\ 0, & w \le (1+\sqrt{\beta})\sqrt{n}\sigma \end{cases}.$$

- Knowledge of r not required
- Noise level σ estimated by comparing last $(m-r)$ s.v. with M-P distribution

$$\hat{\sigma} = \frac{w_{\text{med}}}{\sqrt{n \cdot z_{\text{med}}(\beta)}}$$

Hankel noise structure

- When Z is also Hankel, no analytical solution for μ_Z
- Data-driven singular value shrinkage algorithm (Nadakuditi 2014)
  - Consistent estimate of μ_Z from last $(m-r)$ s.v.

$$\eta_{\text{DD}}(w_i) = \begin{cases} \eta(w_i; \hat{\mu}_Z(w_{r+1}, \ldots, w_m)), & i = 1, \ldots, r \\ 0, & i = r+1, \ldots, m \end{cases}.$$

- Knowledge of r required to distinguish purely noisy s.v.
  - Can be replaced by an upper bound of r

Combining SLRA with optimal shrinkage denoising

Iterative low-rank Hankel matrix denoising

1: **Input:** W, Π, r, ϵ.
2: $W_1 \leftarrow W$
3: **repeat**
4: $\quad W_2 \leftarrow \sum_{i=1}^{r} \eta(w_i; \hat{\mu}_Z(w_{r+1}, \ldots, w_m))\mathbf{u}_i\mathbf{v}_i^{\mathsf{T}} + W_1(\mathsf{I}_n - \Pi)$,
5: $\quad W_1 \leftarrow \mathcal{H}(W_2)$
6: **until** $\|W_1 - W_2\| < \epsilon \|W_1\|$
7: **Output:** $\hat{X} = W_1$.

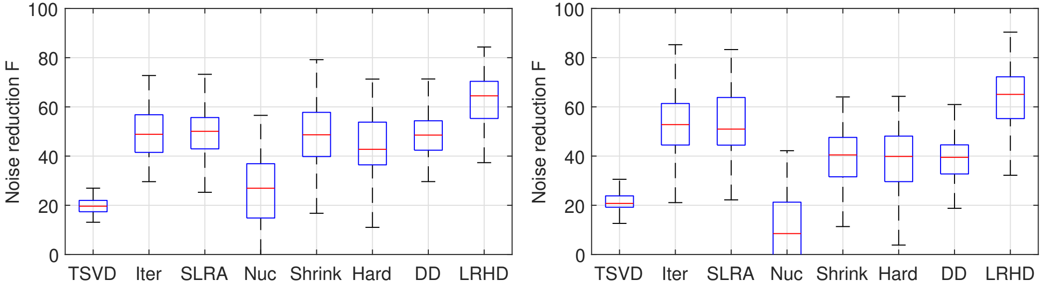
Numerical simulation

- Random fourth-order LTI systems ($r = 4$)
- Zero-mean i.i.d. Gaussian noise in output measurements
- r and σ^2 assumed known if needed
- Performance assessed by noise reduction measure

$$F = 100 \cdot \left(1 - \frac{\left\|X - \hat{X}\right\|_F}{\|X - W\|_F}\right)$$

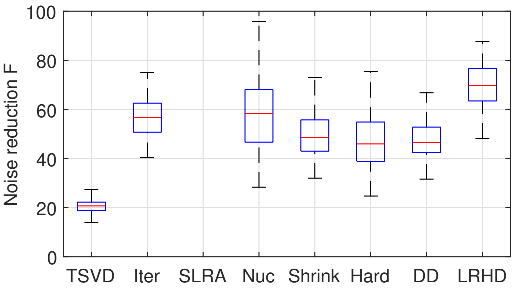
Compared methods

- Truncated singular value decomposition (*TSVD*)
- Structured low-rank approximation methods
  - SLRA by iteration (*Iter*)
  - SLRA by local optimization (*SLRA*)
  - Nuclear norm regularization (*Nuc*)
- Unstructured matrix denoising methods
  - Optimal shrinkage law (*Shrink*)
  - Optimal hard thresholding (*Hard*)
  - Data-driven shrinkage law (*DD*)
- Iterative low-rank Hankel matrix denoising (*LRHD*)

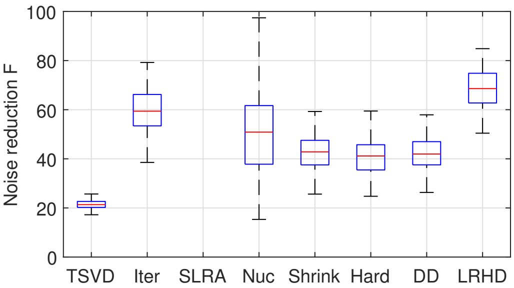


(a) $X \in \mathbb{R}^{8\times 33}$, $\sigma^2 = 0.01$ (b) $X \in \mathbb{R}^{8\times 33}$, $\sigma^2 = 0.001$

Figure: Noise reduction performance for impulse response denoising.



(a) $X \in \mathbb{R}^{8\times 89}$, $\sigma^2 = 0.1$



(b) $X \in \mathbb{R}^{8\times 89}$, $\sigma^2 = 0.01$

Figure: Noise reduction performance for input-output trajectory denoising.

ETH zürich

A novel approach to low-rank Hankel matrix denoising

- Denoising is different from approximation
- Hankel structure enforced by data-driven singular value shrinkage & iterative structural approximation

AUTOMATIC CONTROL LABORATORY

Mingzhou Yin
myin@ethz.ch