Low-rank matrix regression via least-angle regression

Mingzhou Yin, Matthias A. Müller

Institute of Automatic Control
Leibniz University Hannover

Conference on Decision and Control
CDC2025
Rio de Janeiro, Brazil. Dec. 09 - 12

December 10, 2025

zukunft.
niedersachsen

Low-rank matrix regression

- Universal problem for extracting low-order structure from data
- Estimate an unknown low-rank matrix $X \in \mathbb{R}^{m \times n}$ from a noisy measurement Y
- Basic problem: $Y = X + E,\ \mathrm{rank}\,(X) = r,\ r < \min(m, n)$
- Idea: find the best rank-r fit:

$$\hat{X} = \underset{X \in \mathbb{R}^{m \times n}}{\arg\min} \ \|Y - X\|_F^2 \qquad \text{s.t.} \quad \mathrm{rank}\,(X) = r$$

- Closed-form solution: (truncated SVD) $\hat{X} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$
- Virtually anything beyond that is NP-hard
  - Linear instead of direct measurement: $Y = \Phi X + E$, $\Phi \in \mathbb{R}^{p \times m}$
  - X is structured: $X \in \mathbb{M}^{m \times n} \neq \mathbb{R}^{m \times n}$

Motivating examples

$$\hat{X} = \underset{X \in \mathbb{M}^{m \times n}}{\arg\min} \; \|Y - \Phi X\|_F^2 \qquad \text{s.t.} \quad \operatorname{rank}(X) = r$$

- Low-rank network modeling: $x_{k+1} = B^\top x_k + e_k$, transition matrix B is low rank

$$X = B, \quad Y = [x_2 \; \dots \; x_{p+1}]^\top, \quad \Phi = [x_1 \; \dots \; x_p]^\top$$

- System realization: impulse response model of LTI systems $G(q) = \sum_{i=1}^{\infty} g_i q^{-i}$

$$X = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \\ g_2 & g_3 & \cdots & g_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ g_m & g_{m+1} & \cdots & g_{m+n-1} \end{bmatrix}, \qquad \mathbb{M}: \text{ Hankel matrices}$$

An approximate solution

- Replace rank function with tightest convex surrogate: nuclear norm regularization

$$\hat{X}_{\text{nuc}} = \operatorname*{arg\,min}_{X \in \mathbb{M}^{m \times n}} \frac{1}{2} \|Y - \Phi X\|_F^2 + \lambda \|X\|_*$$

- Problems: 1) SDP problems scale unfavorably with the problem size
  2) Rank of $\hat{X}_{\text{nuc}}$ requires tuning λ by trial and error

A different but related problem: sparse learning

- Consider linear regression model

$$Y = \sum_{i=1}^{n_\sigma} \sigma_i \tilde{X}_i + E$$

- Estimate a sparse σ with r nonzero elements $(\mathrm{card}(\sigma) = r)$ from Y and $\left(\tilde{X}_i\right)_{i=1}^{n_\sigma}$
- Idea: Find the best cardinality-r fit:

$$\hat{\sigma} = \underset{\sigma \in \mathbb{R}^{n_\sigma}}{\arg\min} \left\| Y - \sum_{i=1}^{n_\sigma} \sigma_i \tilde{X}_i \right\|^2 \qquad \text{s.t.} \quad \mathrm{card}(\sigma) = r,$$

- NP hard $\rightarrow$ replace cardinality function with tightest convex surrogate: l_1-norm regularization

$$\hat{\sigma}_{l_1}(\lambda) = \underset{\sigma \in \mathbb{R}^{n_\sigma}}{\arg\min} \left\| Y - \sum_{i=1}^{n_\sigma} \sigma_i \tilde{X}_i \right\|_2^2 + \lambda \left\| \sigma \right\|_1$$

- Problem: cardinality of $\hat{\sigma}_{l_1}$ requires tuning λ by trial and error

Least angle regression (LAR)

- An iterative algorithm that obtains $\hat{\sigma}$ for all choices of r
- Assume: $Y, \tilde{X}_i \in \mathbb{R}^p$: vector-valued, $\tilde{X}_i$: normalized with $\left\|\tilde{X}_i\right\|_2 = 1$

1: Find the most output-correlated covariate:

$$i_1 = \arg\max_i |\tilde{X}_i^\top Y|$$

2: Take direction $\zeta_1 = \tilde{X}_{i_1}$ with a step size of η_1 until there exists i_2 that correlates with model residual as much as $\tilde{X}_{i_1}$:

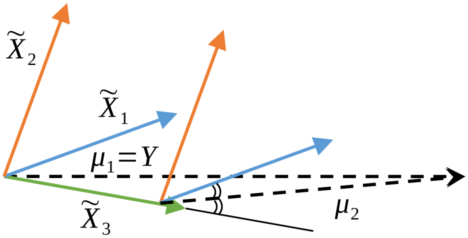
$$(i_2, \eta_1) = \arg\min_{i,\eta} |\eta|$$

$$\text{s.t. } |\tilde{X}_i^\top (Y - \eta \tilde{X}_{i_1})| = |\tilde{X}_{i_1}^\top (Y - \eta \tilde{X}_{i_1})|$$

3: Take bisecting direction:

$$\zeta_2 = \arg\min_\zeta \ \|\zeta\|_2 \text{ s.t. } \tilde{X}_{i_1}^\top \zeta = \tilde{X}_{i_2}^\top \zeta = 1,$$

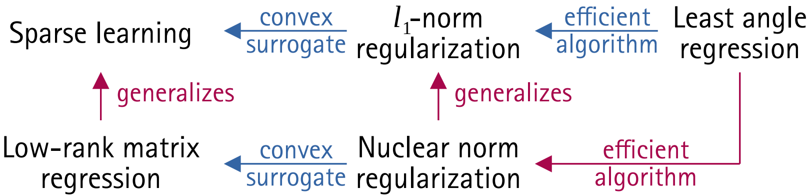
until $\tilde{X}_{i_3}$ comes in with the same residual correlation



$p = 2, \quad n_\sigma = 3$

- LAR solution $\hat{Y}(r) = \sum_{i=1}^r \eta_i \zeta_i$ corresponds to the smallest λ s.t. $\hat{\sigma}_{l_1}(\lambda)$ has cardinality r

Contributions



- This work: by considering low-rank matrix regression as an infinite-dimensional sparse learning problem, develop the LAR algorithm for low-rank matrix regression of unstructured and Hankel matrices

Low-rank matrix regression as sparse rank-1 matrix decomposition

- Idea: rank-r matrix regression = finding a decomposition of r independent rank-1 matrices

$$
\begin{aligned}
\hat{X} = \underset{X \in \mathbb{M}^{m\times n}}{\arg\min}\ & \|Y - \Phi X\|_F^2 \\
\text{s.t.}\ & \operatorname{rank}(X) = r
\end{aligned}
\quad \Longleftrightarrow \quad
\begin{aligned}
\hat{X} &= \sum_{i=1}^{r} \hat{\sigma}_i \hat{X}_i \\
\left(\hat{X}_i, \hat{\sigma}_i\right) &= \underset{X_i, \sigma_i}{\arg\min}\ \left\|Y - \Phi \textstyle\sum_{i=1}^{r} \sigma_i X_i\right\|_F^2 \\
\text{s.t.}\ & X_i = \mathbf{u}_i \mathbf{v}_i^\top,\ i = 1, \ldots, r, \\
& \operatorname{rank}\left(\left[\mathbf{u}_1 \ \cdots \ \mathbf{u}_r\right]\right) = r, \\
& \operatorname{rank}\left(\left[\mathbf{v}_1 \ \cdots \ \mathbf{v}_r\right]\right) = r, \\
& \textstyle\sum_{i=1}^{r} \sigma_i X_i \in \mathbb{M}^{m\times n}
\end{aligned}
$$

- Suppose for all $X \in \mathbb{M}^{m\times n}$, $\operatorname{rank}(X) = r$, there exists a decomposition $X = \sum_{i=1}^{r} \sigma_i X_i$, where $X_i \in \mathbb{M}^{m\times n}$ satisfies the blue constraints
- RHS can be viewed as a sparse learning problem with an infinite set of rank-1 basis $\tilde{X}_i = \Phi X_i$

$$
\hat{\sigma} = \underset{\sigma}{\arg\min} \left\|Y - \textstyle\sum_i \sigma_i \tilde{X}_i\right\|^2 \qquad \text{s.t.} \quad \operatorname{card}(\sigma) = r
$$

- Least angle regression can be applied using a suitable decomposition of X

LAR for unstructured matrix

- Consider the set of orthonormal rank-1 matrices

$$\mathcal{X}_i = \left\{ \mathbf{u}_i \mathbf{v}_i^\top \middle| \; \|\mathbf{u}_i\|_2 = \|\mathbf{v}_i\|_2 = 1, \mathbf{u}_i^\top \mathbf{u}_k = \mathbf{v}_i^\top \mathbf{v}_k = 0, k = 1, \ldots, i-1 \right\}$$

- Normalize Φ by taking SVD: $\Phi = U_\Phi S_\Phi V_\Phi^\top$, $U_\Phi \in \mathbb{R}^{p \times m}$, $S_\Phi, V_\Phi \in \mathbb{R}^{m \times m}$.
- Closed-form solution exists by considering an orthonormal basis of $S_\Phi V_\Phi^\top X_i$

$$\begin{aligned} \hat{X} = \underset{X \in \mathbb{R}^{m \times n}}{\arg\min} \quad & \|Y - \Phi X\|_F^2 \\ \text{s.t.} \quad & \operatorname{rank}(X) = r \end{aligned} \iff \begin{aligned} \hat{X} &= V_\Phi S_\Phi^{-1} \sum_{i=1}^{r} \hat{\sigma}_i^{\mathrm{u}} \hat{X}_i^{\mathrm{u}} \\ \left(\hat{X}_i^{\mathrm{u}}, \hat{\sigma}_i^{\mathrm{u}} \right) &= \underset{X_i^{\mathrm{u}}, \sigma_i^{\mathrm{u}}}{\arg\min} \quad \|Y - U_\Phi \textstyle\sum_{i=1}^{r} \sigma_i^{\mathrm{u}} X_i^{\mathrm{u}}\|_F^2 \\ & \quad \text{s.t.} \qquad X_i^{\mathrm{u}} \in \mathcal{X}_i, \; i = 1, \ldots, r \end{aligned} \qquad (\star)$$

LAR for unstructured matrix: results

Theorem: Closed-form LAR solution for unstructured matrix

Let $U_\Phi^\top Y = U^{\mathrm{u}} S (V^{\mathrm{u}})^\top$ be the SVD of $U_\Phi^\top Y$, where $U^{\mathrm{u}} = [\hat{\mathbf{u}}_1 \ \cdots \ \hat{\mathbf{u}}_m] \in \mathbb{R}^{m\times m}$, $S \in \mathbb{R}^{m\times n}$ with the (i,i)-th element being σ_i, and $V^{\mathrm{u}} = [\hat{\mathbf{v}}_1 \ \cdots \ \hat{\mathbf{v}}_n] \in \mathbb{R}^{n\times n}$.

The LAR solution to the sparse learning problem ($\star$) is given by $\hat{X}_i^{\mathrm{u}} = \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^\top$, $\hat{\sigma}_i^{\mathrm{u}} = \sigma_i - \sigma_{r+1}$.

The low-rank estimate is given by $\hat{X}_{\mathrm{LAR}} = V_\Phi S_\Phi^{-1} \sum_{i=1}^r \left(\sigma_i^0 - \sigma_{r+1}^0\right) \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^\top$.

Corollary: Equivalence to nuclear norm regularization

The closed-form solution is equivalent to the normalized nuclear norm regularization problem:

$$\hat{X}_{\mathrm{nuc,n}}(\lambda) = \operatorname*{arg\,min}_{X \in \mathbb{R}^{m\times n}} \frac{1}{2} \|Y - \Phi X\|_F^2 + \lambda \left\| S_\Phi V_\Phi^\top X \right\|_* ,$$

where $\lambda = \min_\lambda \lambda \ \text{ s.t. } \operatorname{rank}\left(\hat{X}_{\mathrm{nuc,n}}(\lambda)\right) = r$.

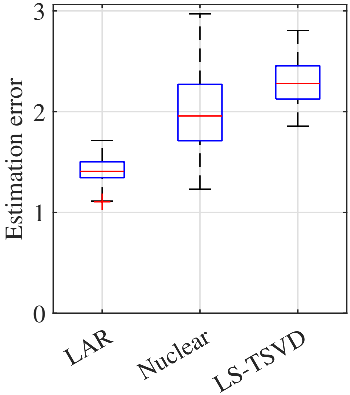
Institute of Automatic Control
Leibniz University Hannover

Numerical results: low-rank network modeling

$$x_{k+1} = B^\top x_k + e_k$$

- 120 Monte Carlo simulations
- $m = n = 40,\ p = 80,\ r = 10$
- Random rank-r B with spectral radius of 0.95
- Gaussian error with $\sigma = 0.01$

- LAR: proposed closed-form solution
- Nuclear: nuclear norm regularization (20-point λ-grid in $[0.01, 0.1]$)
- LS-TSVD: successive least-squares estimation and truncated SVD



Institute of Automatic Control
Leibniz University Hannover

LAR for Hankel matrix

- Orthonormal decomposition no longer possible, rank-1 Hankel decomposition required
- All rank-1 (complex) Hankel matrices can be expressed as

$$X_z = \mathbf{u}_z \mathbf{v}_z^\top,\ \mathbf{u}_z = \begin{bmatrix} 1 & z & \dots & z^{m-1} \end{bmatrix}^\top,\ \mathbf{v}_z = \begin{bmatrix} 1 & z & \dots & z^{n-1} \end{bmatrix}^\top,\ z \in \mathbb{C}$$

- We restrict to Hankel matrices that can be decomposed as $X = \sum_{i=1}^{r} \sigma_{z_i} X_{z_i}$
  - For the system realization example, this corresponds to no repeated poles

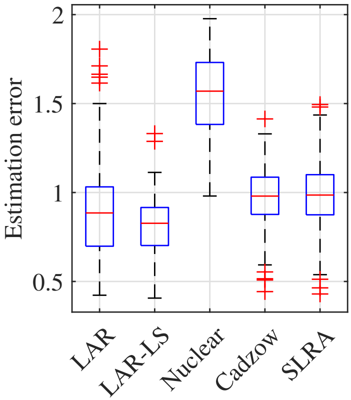
$$\begin{aligned} \hat{X} = \underset{X \in \mathbb{H}^{m\times n}}{\arg\min} \quad & \|Y - \Phi X\|_F^2 \\ \text{s.t.} \quad & \operatorname{rank}(X) = r \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \hat{X} &= \sum_{i=1}^{r} \hat{\sigma}_{z_i} \hat{X}_{z_i} \\ \left(\hat{X}_{z_i}, \hat{\sigma}_{z_i}\right) &= \underset{X_{z_i}^{\psi_i}, \sigma'_{z_i}}{\arg\min} \quad \|Y - \Phi \textstyle\sum_{i=1}^{r} \sigma_{z_i} X_{z_i}\|_F^2 \\ & \text{s.t.} \quad X_z = \mathbf{u}_z \mathbf{v}_z^\top,\ i = 1, \dots, r, \\ & \qquad\quad z_i \neq z_j,\ \forall\, i \neq j \end{aligned}$$

- A real-valued reformulation for complex X_z is presented in the paper, but omitted here for clarity
- No closed-form solution, LAR iterations can be implemented

Numerical results: system realization

- 120 Monte Carlo simulations
- $m = 80,\ n = 20,\ r = 6$
- Sixth-order benchmark system
- Gaussian error with $\sigma = 0.1$

- LAR: proposed LAR algorithm
- LAR-LS: LAR algorithm followed by least-squares projection
- Nuclear: nuclear norm regularization (20-point λ-grid in $[0.1, 1]$)
- Cadzow, SLRA: other existing algorithms



Low-rank matrix regression via least-angle regression

- Connect low-rank matrix regression with sparse learning
- Closed-form LAR solution for unstructured matrices
- Efficient LAR iterations for Hankel matrices

Institute of Automatic Control
Leibniz University Hannover

Mingzhou Yin
yin@irt.uni-hannover.de