Statistical rank selection for incomplete low-rank matrices

Yao Xie

School of Industrial and Systems Engineering
Georgia Institute of Technology

May 15, 2019

Joint work with Alexander Shapiro and Rui Zhang

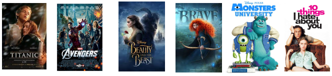
Present at ICASSP 2019

Low-rank matrix completion

- Incomplete and noisy observations

$$Y_{i,j} = X_{i,j} + \epsilon_{i,j}, \quad (i,j) \in \Omega \subset [m_1] \times [m_2]$$

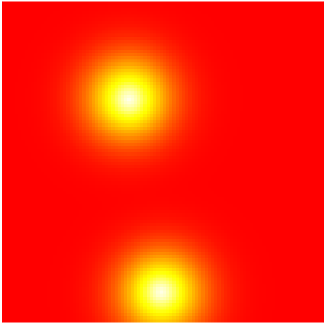
- Recommender systems

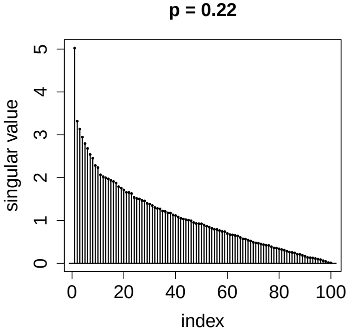


	5			2	
1		3	2		
	5			1	
3		4			3
5				2	
	3	2			1

Example: Determine the number of sources from incomplete observations

$$Y_{i,j} = X_{i,j} + \epsilon_{i,j}, \quad (i,j) \in \Omega \subset [m_1] \times [m_2]$$



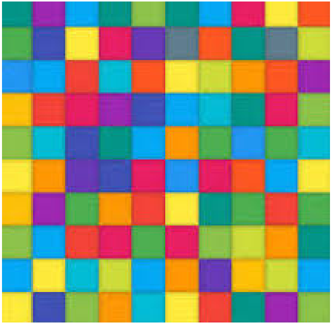


(Chen, Mitra '17)

Low-rank models

$$Y_{i,j} = \underbrace{X_{i,j}}_{\text{low-rank}} + \epsilon_{i,j}, \quad (i,j) \in \Omega \subset [m_1] \times [m_2]$$

low-rank



high-rank

Prior work

- **Convex** relaxation: **noiseless**
  (Candes, Recht '08, Candes, Tao '08, Gross '09)

$$\min_{Z} \|Z\|_* \text{ subject to } Y_{ij} = Z_{ij},\ (i,j) \in \Omega.$$

- **Convex** relaxation robustness to **noise**
  (Candes, Plan '09, Negahban, Wainwright '10, Koltchinskii et al. '10)

$$\min_{Z} \quad \underbrace{f(Z;Y)}_{\text{empirical loss}} \quad + \lambda \|Z\|_*$$

- **Non-convex** optimization
  Burer, Monteiro '03, Rennie Srebro '05, Jain, Netrapalli, Sanghavi '12, Ma, Wang, Chi, Chen '17 ...

$$\min_{U \in \mathbb{R}^{n_1 \times r}, V \in \mathbb{R}^{n_1 \times r}} \sum_{(i,j) \in \Omega} [(UV^T)_{ij} - Y_{ij}]^2 + \text{regularizer}$$

Motivation

- Select "rank" parameter in algorithm

$$\min_{U \in \mathbb{R}^{n_1 \times r}, V \in \mathbb{R}^{n_1 \times r}} \sum_{(i,j) \in \Omega} [\underbrace{(UV^T)_{ij}}_{\text{low-rank } X} - Y_{ij}]^2 + \text{regularizer}$$

- Determine "true" rank when the underlying matrix is low-rank

$$Y_{i,j} = \underbrace{X_{i,j}}_{\text{low-rank}} + \epsilon_{i,j}, \quad (i,j) \in \Omega \subset [m_1] \times [m_2]$$

Problem formulation

Noisy and possibly biased observations of a subset of matrix entries

$$Y_{ij} = X^*_{ij} + N^{-1/2}\Delta_{ij} + \varepsilon_{ij},\ (i,j) \in \Omega,$$

- $X^* \in \mathcal{M}_{r^*}$ low-rank matrix
- N effective sample size
- Δ_{ij} deterministic bias term
- $N^{1/2}\varepsilon_{ij} \xrightarrow{\text{in dist}} \mathcal{N}(0, \sigma^2_{ij})$ variance can be different

Goal: determine r^* using statistical test procedure

Assumptions

Typical assumptions

- Non-adaptive, random sampling: each $(i, j) \in \Omega$ independently with probability p
- Random noise: i.i.d. sub-Gaussian noise
- Ground truth: M^* is low-rank

Here

- Ω is deterministic

Our contribution

- Develop a new statistical test procedure to determine the rank
- Solve a sequence of "fitting" problems with different $\color{red}{r}$

$$\min_{U \in \mathbb{R}^{n_1 \times \color{red}{r}}, V \in \mathbb{R}^{n_1 \times \color{red}{r}}} \sum_{(i,j) \in \Omega} [(UV^T)_{ij} - Y_{ij}]^2$$

- Examine residuals to decide
  Example: true rank is 6.

Table: *sequential rank test*

rank	p-value	$\hat{\sigma}^2 (= \bar{Z})$	rank	p-value	$\hat{\sigma}^2 (= \bar{Z})$
1	0.82	34995.5	5	0.84	5050.63
2	0.86	26751.3	**6**	**0.43**	**97.7**
3	0.92	18719.6	7	0.76	96.6
4	0.62	11231.8	8	0.96	96.7

Formal results: How to select r?

- Solve a sequence of weighted least squares test statistic

$$T_N(r) := N \min_{Y \in \mathcal{M}_r} \sum_{(i,j)\in\Omega} w_{ij} \left(M_{ij} - Y_{ij}\right)^2 ,$$

  $w_{ij} := 1/\hat{\sigma}_{ij}^2$ with $\hat{\sigma}_{ij}^2$ being consistent estimates of σ_{ij}^2

  $\mathcal{M}_r$: (manifold) of all rank-r matrices
- m $= |\Omega|$ number of measurements

Asymptotic properties of test statistic

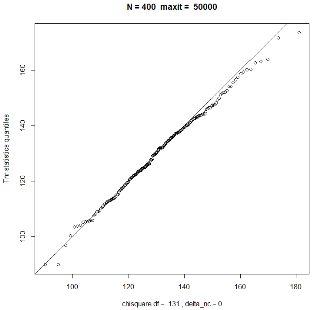
$$T_N(r) \Rightarrow \chi^2(df_r, \delta_r),$$

1. degrees of freedom

$$\mathrm{df}_r = |\Omega| - \dim(\mathcal{M}_r) = \mathsf{m} - r(n_1 + n_2 - r),$$

2. noncentrality parameter

$$\delta_r = \min_{H \in \mathcal{T}_{\mathcal{M}_r}(Y^*)} \sum_{(i,j)\in\Omega} \sigma_{ij}^{-2} \left(\Delta_{ij} - H_{ij}\right)^2 .$$



Sequential test procedures

- Sequentially test $r = 1, 2, 3, \ldots$ using $T_N(r)$
- "null" hypothesis that the "true" rank is r^*
- null hypothesis is rejected if

  $T_N(r)$ is **large** enough on the scale of the χ^2 distribution

- perform such tests sequentially for increasing values of r

Table: *sequential rank test*

rank	p-value	$\hat{\sigma}^2 (= \bar{Z})$	rank	p-value	$\hat{\sigma}^2 (= \bar{Z})$
1	0.82	34995.5	5	0.84	5050.63
2	0.86	26751.3	**6**	**0.43**	**97.7**
3	0.92	18719.6	7	0.76	96.6
4	0.62	11231.8	8	0.96	96.7

Additional comments

- Role of values Δ_{ij}: suggest that "true" model is true only approximately
- noncentrality parameter

$$\delta_r = \min_{H \in \mathcal{T}_{\mathcal{M}_r}(Y^*)} \sum_{(i,j)\in\Omega} \sigma_{ij}^{-2} \left(\Delta_{ij} - H_{ij}\right)^2 .$$

indicates the deviation from the exact rank r model.

“Single” matrix observation

- Suppose $N = 1$, $\Delta_{ij} = 0$ and $\varepsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.
- Consider a sequence of index set

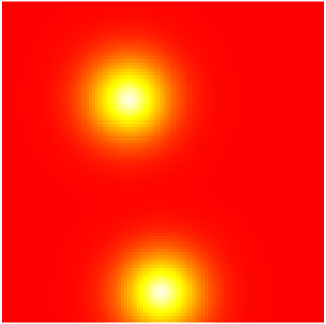
$$\Omega_0 \supset \Omega_1 \supset \Omega_2 \supset \cdots \supset \Omega_K,$$

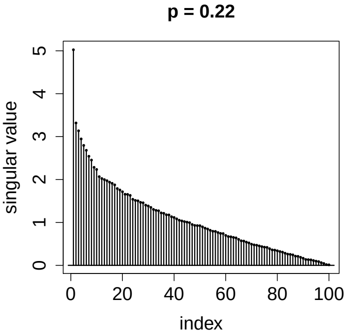
  $|\Omega_{k-1}| - |\Omega_k| = L$, $\forall k = 1 \cdots K$.
- Let

$$X_i = \min_{Y \in \mathcal{M}_r} \sum_{(i,j) \in \Omega_i} (M_{ij} - Y_{ij})^2,$$
$$Z_i = (X_{i-1} - X_i)/L.$$

- $\sqrt{K}(\bar{Z} - \sigma^2)$ converge in distribution to $\mathcal{N}(0, 2\sigma^4/L)$

Example: Determine the number of sources





rank	p-value	rank	p-value
1	0.00	4	1.00
2	**0.15**	5	1.00
3	0.98	6	1.00

Summary

$$Y_{ij} = X^*_{ij} + N^{-1/2}\Delta_{ij} + \varepsilon_{ij},\ (i,j) \in \Omega,$$

- How to select rank r? **Sequential χ^2 test**
- Test statistic

$$T_N(r) := N \min_{Y \in \mathcal{M}_r} \sum_{(i,j)\in\Omega} w_{ij} \left(M_{ij} - Y_{ij}\right)^2 \Rightarrow \chi^2(df_r, \delta_r),$$

rank	p-value	rank	p-value
1	0.00	4	1.00
2	**0.15**	5	1.00
3	0.98	6	1.00

- Role of values Δ_{ij}: suggest that "true" model is true only approximately; non-central parameter δ_r indicates so

Thank you!

References

1. Matrix completion with deterministic pattern - a geometric perspective. A. Shapiro, Y. Xie, and R. Zhang. IEEE Transactions on Signal Processing. Volume: 67 , Issue: 4 , Feb.15, 15 2019