Matrix Completion with Model-free Weighting

Jiayi Wang (*Texas A&M University*), **Raymond K. W. Wong** (*Texas A&M University*),
Xiaojun Mao (*Fudan University*) and Kwun Chuen Gary Chan (University of Washington)

ICML 2021

Background

- Matrix completion
  - to complete a high-dimensional matrix (often low-rank) from its partial and possibly noisy observation

- Existing work under non-uniform missingness is relatively sparse
  1. No adjustment: a form of robustness result for *uniform* empirical risk minimization with regularization
  2. Active adjustment: via a model of missingness (e.g., rank-1)
     $\rightsquigarrow$ how to choose/estimate the model?

- Our work — active adjustment via balancing weights
  - actively adjusts for the non-uniform missingness, without explicitly modeling the probabilities of observation

Setup

- Target matrix $\boldsymbol{A}_\star = (A_{\star,ij})_{i,j=1}^{n_1,n_2}$

- Contaminated target matrix $\boldsymbol{Y} = (Y_{ij})_{i,j=1}^{n_1,n_2}$

$$Y_{ij} = A_{\star,ij} + \epsilon_{ij}, \quad i = 1, \dots, n_1; j = 1, \dots, n_2,$$

where $\{\epsilon_{ij}\}$ are independent errors with zero mean.

- Observation indicator matrix $\boldsymbol{T} = (T_{ij})_{i,j=1}^{n_1,n_2} \in \mathbb{R}^{n_1 \times n_2}$

$$T_{ij} = \begin{cases} 1, & \text{if } Y_{ij} \text{ is observed} \\ 0, & \text{otherwise} \end{cases}$$

where $\{T_{ij}\}$ are independent Bernoulli random variables with $\pi_{ij} = \Pr(T_{ij} = 1)$.

Motivation

- A common strategy: weighted empirical risk

$$\widehat{R}_W(\boldsymbol{A}) = \frac{1}{n_1 n_2} \| \boldsymbol{T} \circ \boldsymbol{W}^{\circ(1/2)} \circ (\boldsymbol{Y} - \boldsymbol{A}) \|_F^2$$

- A natural choice of $\boldsymbol{W}$: inverse probability ($W_{ij} = 1/\pi_{ij}$)
$\rightsquigarrow$ unknown in practice; high-dimensional in nature; unstable estimations due to extreme weights

Motivation

- A novel balancing idea:

$$\frac{1}{n_1 n_2}\|\boldsymbol{T} \circ \boldsymbol{W}^{\circ(1/2)} \circ (\boldsymbol{A}_\star - \boldsymbol{A})\|_F^2 \approx \frac{1}{n_1 n_2}\|\boldsymbol{A}_\star - \boldsymbol{A}\|_F^2,$$

$$0 \approx \frac{1}{n_1 n_2}\left|\langle(\boldsymbol{T} \circ \boldsymbol{W} - \boldsymbol{J}) \circ \boldsymbol{\Delta}, \boldsymbol{\Delta}\rangle\right|, \qquad \boldsymbol{\Delta} = \boldsymbol{A}_\star - \boldsymbol{A},$$

where $\boldsymbol{J}$ is a matrix of ones

Motivation

- Find weights $\boldsymbol{W}$ that minimize the uniform balancing error

$$\sup_{\boldsymbol{\Delta}\in\mathcal{D}_{n_1,n_2}} S(\boldsymbol{W},\boldsymbol{\Delta}) := \sup_{\boldsymbol{\Delta}\in\mathcal{D}_{n_1,n_2}} \frac{1}{n_1 n_2} \left|\langle(\boldsymbol{T}\circ\boldsymbol{W}-\boldsymbol{J})\circ\boldsymbol{\Delta},\boldsymbol{\Delta}\rangle\right|,$$

  for a (standardized) set $\mathcal{D}_{n_1,n_2}$, induced by the hypothesis class $\mathcal{A}_{n_1,n_2}$ of $\boldsymbol{A}_\star$.

- Additional consideration to the choice of $\mathcal{A}_{n_1,n_2}$: computation of the uniform balancing error

Relaxation

Lemma 1

For any matrices $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n_1 \times n_2}$, we have

$$|\langle \boldsymbol{C} \circ \boldsymbol{B}, \boldsymbol{B} \rangle| \leq \|\boldsymbol{C}\| \|\boldsymbol{B}\|_{\max} \|\boldsymbol{B}\|_* \leq \sqrt{n_1 n_2} \|\boldsymbol{C}\| \|\boldsymbol{B}\|_{\max}^2.$$

- We have
$$S(\boldsymbol{W}, \Delta) \leq \sqrt{n_1 n_2} \|\boldsymbol{T} \circ \boldsymbol{W} - \boldsymbol{J}\| \|\Delta\|_{\max}^2$$
- Choose $\mathcal{A}_{n_1,n_2} = \{\boldsymbol{A} : \|\boldsymbol{A}\|_{\max} \leq \beta\}$ and then $\mathcal{D}_{n_1,n_2} = \{\boldsymbol{\Delta} : \|\boldsymbol{\Delta}\|_{\max} \leq 2\beta\}$

Novel weights

- The proposed weights:

$$\widehat{\boldsymbol{W}} = \arg\min_{\boldsymbol{W}} \|\boldsymbol{T} \circ \boldsymbol{W} - \boldsymbol{J}\|$$
$$\text{subject to} \quad \|\boldsymbol{T} \circ \boldsymbol{W}\|_F \leq \kappa \quad \text{and} \quad W_{ij} \geq 1,$$

  where $\kappa \geq \sum_{i,j} T_{ij}$ is a tuning parameter

- Optimization: convex; analytic form of the subgradient is obtainable for the dual Lagrangian form

- Theoretical guarantee of balancing: a non-asymptotic upper bound for the uniform balancing error $\sup_{\|\boldsymbol{\Delta}\|_{\max} \leq \beta'} S(\widehat{\boldsymbol{W}}, \boldsymbol{\Delta})$ (see Theorem 1)

Estimation of $\boldsymbol{A}_\star$

- A hybrid constraint/regularization:
  1. Max-norm constraint: from the construction of $\mathcal{A}_{n_1,n_2}$
  2. Nuclear-norm regularization: sometimes produces tighter relaxation; shows benefits in exact low-rank cases

- Hybrid weighted estimator:

$$\widehat{\boldsymbol{A}} = \underset{\|\boldsymbol{A}\|_{\max} \le \beta}{\arg\min} \left\{ \widehat{R}_{\widehat{\boldsymbol{W}}}(\boldsymbol{A}) + \mu \|\boldsymbol{A}\|_* \right\},$$

  where $\|\cdot\|_*$ denotes the nuclear norm, and $\beta > 0$, $\mu \ge 0$ are tunning parameters

- Optimization:
  1. Convex (original formulation): ADMM algorithm
  2. Nonconvex (via a nonconvex formulation): projected gradient descent algorithm

Theoretical guarantees

- Theoretical guarantee of recovery: a non-asymptotic upper bound for $(n_1 n_2)^{-1}\|\widehat{\boldsymbol{A}} - \boldsymbol{A}_\star\|_F^2$ (see Theorem 2)
- Two asymptotic regimes:
  - asymptotically homogeneous: $\pi_L \asymp \pi_U$ (common asymptotic framework)
  - asymptotically heterogeneous: $\pi_L = o(\pi_U)$ (a good model for highly varying probabilities; empirical evidence from Mao et al. (2020))

  where $\pi_L = \min \pi_{ij}$, $\pi_U = \max \pi_{ij}$
- See Section 5 for the comparison with existing work under asymptotically homogeneous settings
- For asymptotically heterogeneous settings, our bound scales with $\pi_L^{-1/2}$, which is significantly better than the existing upper bounds $(\pi_L^{-1}\pi_U^{1/2})$
- A new minimax lower bound: the scaling $\pi_L^{-1/2}$ cannot be improved (see Theorem 3)

Acknowledgements

- US National Science Foundation[1]
- National Natural Science Foundation of China
- Shanghai Sailing Program
- Science and Technology Commission of Shanghai Municipality
- High performance research computing resources provided by Texas A&M University

[1] Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.