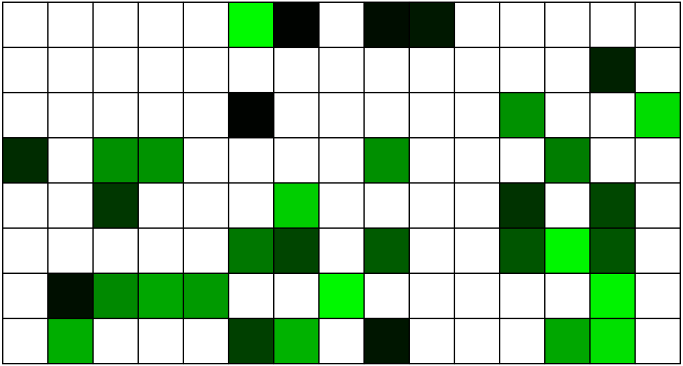
An algorithm for two-cost budgeted matrix completion

Dong Hu

Rensselaer Polytechnic Institute, Troy, NY

Problem

$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a low-rank (rank-r) matrix from which we have noisily observed only a few entries $\boldsymbol{A}_{\text{obs}}$, with indices Ω_e.



$\boldsymbol{A}_{\text{obs}}$

Recover a good approximation $\overline{\boldsymbol{A}}$ of $\boldsymbol{A}$.

Applications

Ubiquitous in statistics, applied math, electrical engineering.

- Recommender systems
- Genomics
- Multi-task learning
- Computer vision
- ... many, *many* more

Classical Approach: Nuclear Norm Completion

$$\overline{\boldsymbol{A}} = \arg\min_{\boldsymbol{Z}} \operatorname{rank}(\boldsymbol{Z}) \quad \text{s.t. } \|\mathcal{P}_{\Omega_e}(\boldsymbol{Z}) - \mathcal{P}_{\Omega_e}(\boldsymbol{A}_{\text{obs}})\|_F \leq \delta$$

- $\mathcal{P}_{\Omega_e}(\cdot)$ zeroes out all unobserved entries.
- The parameter δ is chosen to correspond to the noise level
- This problem is nonconvex, and NP-hard

$$\overline{\boldsymbol{A}} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_\star \quad \text{s.t. } \|\mathcal{P}_{\Omega_e}(\boldsymbol{Z}) - \mathcal{P}_{\Omega_e}(\boldsymbol{A}_{\text{obs}})\|_F \leq \delta$$

- $\|\cdot\|_\star$ is the nuclear norm, a convex proxy for the rank
- This problem is convex
- If Ω_e are sampled i.i.d. uniformly at random, solution has approximation guarantees

Drawbacks of Nuclear Norm MC

The classical uniform sampling model is restrictive:

- It requires *incoherence*: all the entries of the matrix are equally important.
- It cannot take advantage of multiple sampling modalities with different cost-vs-accuracy tradeoffs.
- It requires that $O((n+m)r\log(n+m))$ entries of the matrix can be observed with high precision to obtain approximation guarantees. Cannot handle budgets for observations.

Contributions

This project:

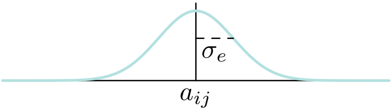
- Investigate a two-cost budgeted matrix completion framework.
- Test a regression-based algorithm for this budgeted MC problem. It allows $\boldsymbol{A}$ to be row-incoherent, and exploits the cost-vs-accuracy tradeoffs of the two sampling modalities.
- Empirically validate the superior performance of this algorithm for two-cost budgeted MC.

Two-cost budgeted completion

An idealized model corresponding to two sampling models with different cost-vs-accuracy tradeoffs. The experimentalist:

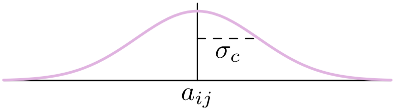
- Has a finite budget $B > 0$.
- At cost p_e can draw a single low-noise entry observation,

$$(\boldsymbol{A}_{\text{obs}})_{ij} \sim \mathcal{N}(\boldsymbol{A}_{ij}, \sigma_e^2)$$



- At cost p_c can observe an entire column with higher noise,

$$(\boldsymbol{A}_{\text{obs}})_{:,j} \sim \mathcal{N}(\boldsymbol{A}_{:,j}, \sigma_c^2)$$



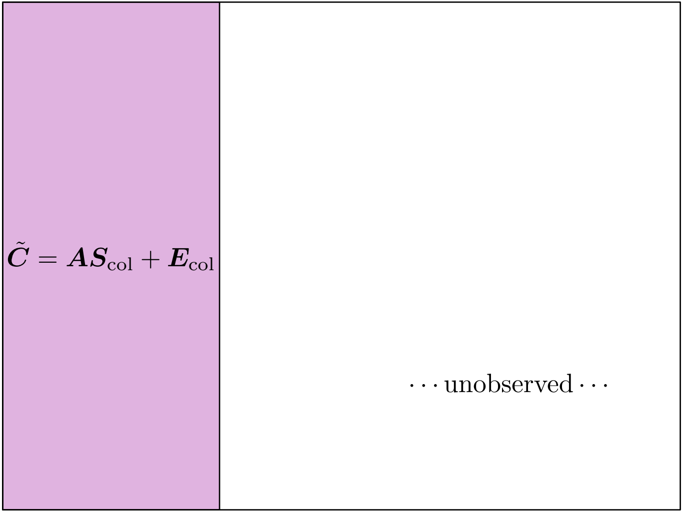
- $p_c \ll p_e m$: the amortized cost of column sampling is much lower than entry sampling.
- $\sigma_c^2 > \sigma_e^2$: column noise is higher than entrywise noise.

How to allocate the budget between entry and column observations to obtain an accurate $\overline{\boldsymbol{A}}$?

- Let d be the number of column observations.
- Classical MC model chooses $d = 0$ and spends all of the budget on high-fidelity entry observations.
- In the low-budget case, $B \leq (n+m)r\log(n+m)p_e$, this is not enough to sample the entries needed to get recovery guarantees.
- **Hypothesis: one can use a mix of low-fidelity column observations with high-fidelity entrywise observations to get recovery guarantees even in the low-budget case.**

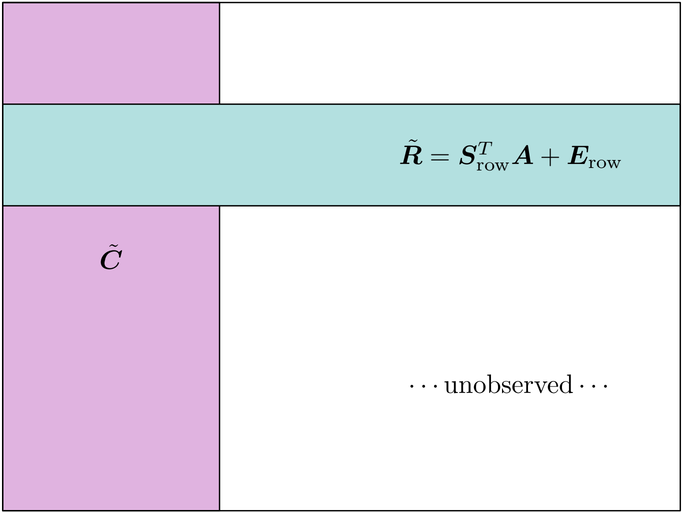
Algorithm

Sample d noisy columns $\tilde{\boldsymbol{C}}$ to capture the column span of $\boldsymbol{A}$.



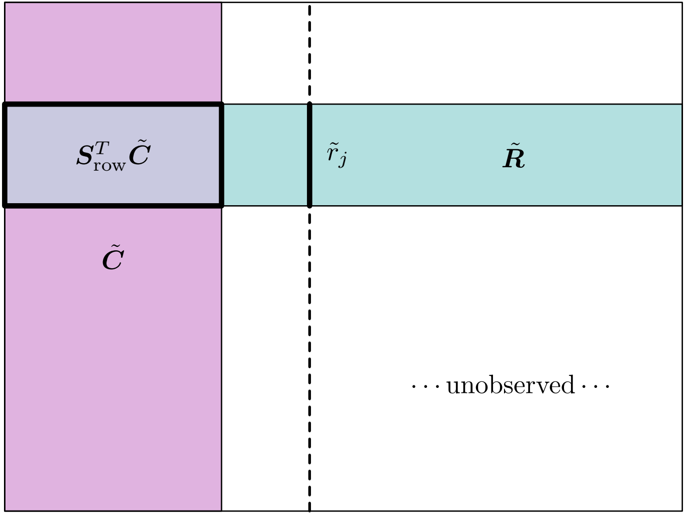
Algorithm

Sample s noisy rows (entrywise) $\tilde{\boldsymbol{R}}$ from all columns of $\boldsymbol{A}$.



Algorithm

Regress observations against $\tilde{\boldsymbol{C}}$, so $\overline{\mathbf{a}}_j = \tilde{\boldsymbol{C}}(\boldsymbol{S}_{\text{row}}^T\tilde{\boldsymbol{C}})^\dagger\tilde{\mathbf{r}}_j$.



Algorithm: the details

- The number of column samples d depends on the rank, noise level, and conditioning of the matrix $\boldsymbol{A}$.
- The number of rows $s = \tilde{O}(d)$.
- Use ridge regression to better handle noise, so this Algorithm's estimate is

$$\overline{\boldsymbol{A}} = \tilde{\boldsymbol{C}}\boldsymbol{X},$$

where

$$\boldsymbol{X} = \arg\min_{\boldsymbol{Z}} \|\tilde{\boldsymbol{R}} - (\boldsymbol{S}_{row}^T \tilde{\boldsymbol{C}})\boldsymbol{Z}\|_F^2 + \lambda \|\boldsymbol{Z}\|_F^2.$$

Empirical results: Baselines

The performance of this Algorithm is compared against that of four baselines: CUR+, a regression-based baseline, and three nuclear norm-based baselines

1) CUR+: a regression algorithm originally designed for noiseless matrix completion, is adapted to the two-cost budgeted noisy setting in a straightfoward manner. Given a value of d, low-precision observations of $d/2$ noisy columns $\tilde{\boldsymbol{C}}$ and $d/2$ noisy rows $\tilde{\boldsymbol{R}}$ are sampled uniformly with replacement from $\boldsymbol{A}$, and the remaining budget is used to sample entries to form Ω_e and $\boldsymbol{A}_{\text{obs}}$. The approximation is $\overline{\boldsymbol{A}} = \tilde{\boldsymbol{C}}\boldsymbol{U}\tilde{\boldsymbol{R}}$, where

$$\boldsymbol{U} = \arg\min_{\boldsymbol{Z}} \|\mathcal{P}_{\Omega_e}(\boldsymbol{A}_{\text{obs}}) - \mathcal{P}_{\Omega_e}(\tilde{\boldsymbol{C}}\boldsymbol{Z}\tilde{\boldsymbol{R}})\|_F^2.$$

2) NNa: Nuclear norm minimization for all uniform samples.

$$\overline{\boldsymbol{A}} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_\star$$
$$\text{s.t.} \, \|\mathcal{P}_{\Omega_e}(\boldsymbol{Z}) - \mathcal{P}_{\Omega_e}(\boldsymbol{A}_{\text{obs}})\|_F \leq \delta$$

All the budget is spent to sample entries Ω_e and form $\boldsymbol{A}_{\text{obs}}$.

3) NNs: Nuclear norm minimization with separate penalizations.

$$\overline{\boldsymbol{A}} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_\star$$
$$\text{s.t.} \, \|\mathcal{P}_{\Omega_e}(\boldsymbol{Z}) - \mathcal{P}_{\Omega_e}(\boldsymbol{A}_{\text{obs}})\|_F \leq \delta_e^2$$
$$\text{s.t.} \, \|\mathcal{P}_{\Omega_c}(\boldsymbol{Z}) - \mathcal{P}_{\Omega_c}(\boldsymbol{A}_{\text{obs}})\|_F \leq \delta_c^2$$

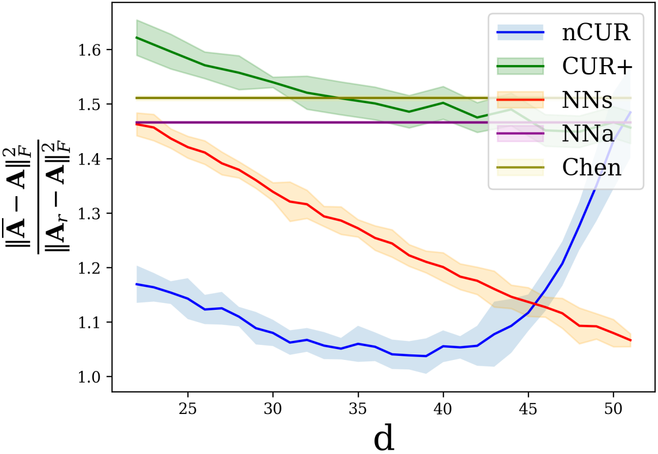
Noisy observations of d columns are uniformly sampled to form Ω_c, the corresponding indices, and placed in $\boldsymbol{A}_{\text{obs}}$, then the remaining budget is spent to uniformly sample entries Ω_e, which are added to $\boldsymbol{A}_{\text{obs}}$.

4) Chen: An adaptation of a two-phase nuclear norm minimization method. A portion of the budget is sampled to sample entries uniformly at random with low noise, which are used to estimate the leverage scores of the rows of $\boldsymbol{A}$. The remaining budget is used to sample low-noise entries Ω_e according to the entrywise leverage scores, rather than uniformly at random, to form $\boldsymbol{A}_{\text{obs}}$. The NNa formulation is used with these samples.

$$\begin{aligned}\overline{\boldsymbol{A}} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_\star \\ \text{s.t.} \|\mathcal{P}_{\Omega_e}(\boldsymbol{Z}) - \mathcal{P}_{\Omega_e}(\boldsymbol{A}_{\text{obs}})\|_F \leq \delta\end{aligned}$$

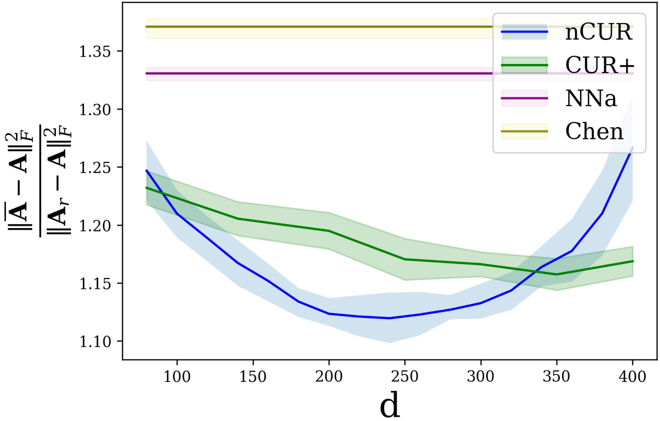
For all of these methods, cross-validation is used to select the hyperparameters. ADMM is used to obtain the nuclear norm estimators.

Jester dataset



$\boldsymbol{A} \in \mathbb{R}^{7200 \times 100}$, joke ratings dataset. Budget chosen so that only 11% of the entries can be observed if all the budget is used. Noise levels: $\sigma_e^2 = 0.04$ and $\sigma_c^2 = 2$. Ratio of costs: $p_c/(mp_e) = 0.2$.

MovieLens dataset



$\boldsymbol{A} \in \mathbb{R}^{1682 \times 943}$, movie ratings dataset. Budget chosen so that only 10.6% of the entries can be observed if all the budget is used. Noise levels: $\sigma_e^2 = 0.003$ and $\sigma_c^2 = 0.06$. Ratio of costs: $p_c/(mp_e) = 0.2$. NNs is not shown as it has orders of magnitude higher error.

Conclusion

- Investigated the two-cost budgeted matrix completion problem
- Tested the algorithm, which is a ridge-regression based algorithm for this problem
- Empirically demonstrated its superior performance to nuclear norm baselines for this problem
- Future question: how to determine an optimal d for a given budget?

Thank you!