

Research on Restoring Clipped Low-rank Matrices

Takeshi Teshima (Graduate School of Frontier Sciences)

July 10th, 2019

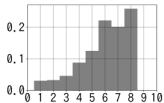
© IRCN seminar

① Problem : Ceiling Effect

= Measure has upper bound.



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



② Matrix Completion

= Restore low-rank matrix.

4	2	3
4	3	4
4	3	4



4	2	
		4
4	3	4

- missing
- noise
etc.

③ So, we worked on...

Clipped Matrix Completion

Low-rank

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4



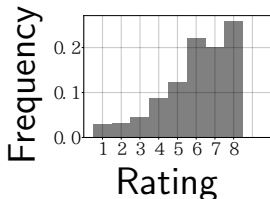
	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Clip (+missing)

Ceiling Effect

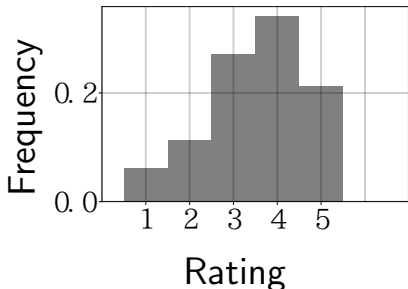
Situation: measurement tool has an upper bound [1]. Exceeding values are observed after clipping to the upper bound.

- Ex. 5-scale response
- Too many “5” \Rightarrow question failed to capture true preferences.

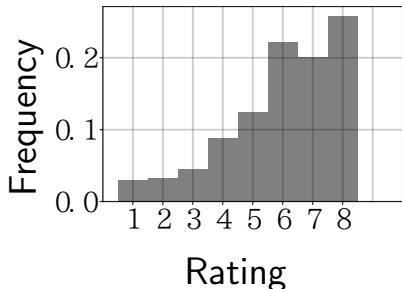


Ceiling effect in machine learning data 3

MovieLens 100K



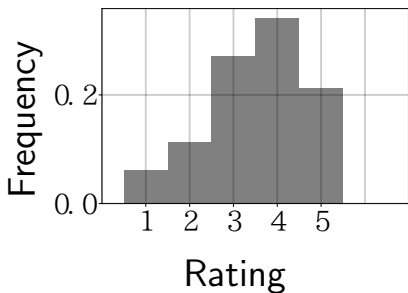
FilmTrust



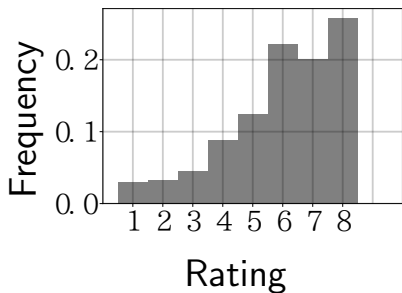
- Benchmark data of recommendation systems.
- Right-truncated shape is typical for ceiling effect

How can we investigate the true values of data that is prone to ceiling effects?

Movielens 100K



FilmTrust

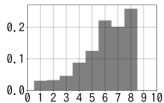


① Problem : Ceiling Effect

= Measure has upper bound.



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



② Matrix Completion

= Restore low-rank matrix.

4	2	3
4	3	4
4	3	4



4	2	
		4
4	3	4

- missing
- noise
etc.

③ So, we worked on...

Clipped Matrix Completion

Low-rank

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

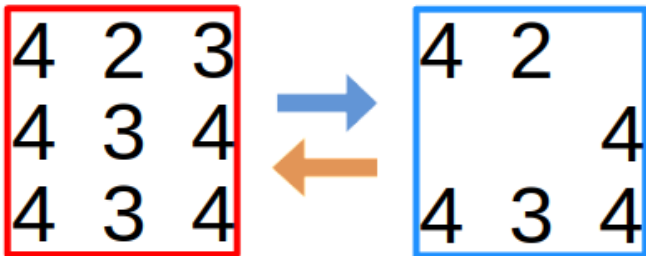


	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Clip (+missing)

② Matrix completion

- A technique to recover a matrix from deficits e.g. missing [3].











- Goal: fill in the blanks
- Missing, noise, quantization, etc. For each deficit, methods are developed.

② Application of matrix completion

7

- Example application: movie recommendation system

Movie

					
User 		5			4
User 	5		1		5
User 	5	1	3		3

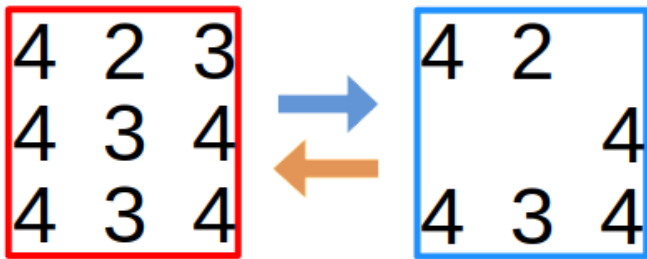
- There are conditions for MC to be possible.
- **Low-rank** and well-behaved matrix \Rightarrow recoverable [3].

The principle of low-rank completion

- Low-rank ... each entry is an inner product of row-/column-vectors.
- How to fill in: estimate feature vectors \rightarrow compute inner products



The principle of low-rank matrix completion: recover matrices from missing etc.

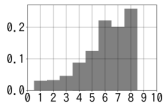


① Problem : Ceiling Effect

= Measure has upper bound.



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



② Matrix Completion

= Restore low-rank matrix.

4	2	3
4	3	4
4	3	4



4	2	
		4
4	3	4

- missing
- noise
etc.

③ So, we worked on...

Clipped Matrix Completion

Low-rank

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4



	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Clip (+missing)

Completing low-rank matrix from its clipped observations

	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Observation



4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

Underlying matrix

Problem (Clipped Matrix Completion; CMC)

Accurately recover the underlying matrix from a random subset of its clipped observations and the known clipping threshold.

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

True values
(unknown)



Obs.

	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Observed data
(clip @ 10)

Rec.



4.0	7.0	4.0	7.0	4.0
-0.0	3.0	6.0	14.9	11.9
4.0	6.0	2.0	2.0	0.0
2.0	6.0	7.0	15.9	11.9
8.0	13.0	6.0	9.0	4.0

After
completion

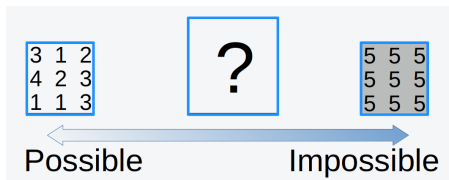
- Proposed problem setting → Clipped matrix completion.
- Recovery possible? → Exact recovery possible under conditions.
- Recovery method? → Minimize squared hinge loss + regularization term
- Experimentally? → Resilience to ceiling effect may improve recommendation systems!

1. Main theorem: when is recovery possible?
2. Proposed method: how to do the recovery?
3. Theoretical guarantee of the proposed method (omitted)
4. Experimental evaluation

Technical detail ①: when is recovery possible?

14

- Motivation for theory: not all clipped matrices can be completed.



- ▶ There are evident cases where recovery is impossible.
- ▶ There are cases where no treatment for clipping is required.
- Main theorem: a sufficient condition for the recovery to be possible.
 - at least, there are cases where recovery is feasible (even with non-negligible clipping).

Assumptions (informal)

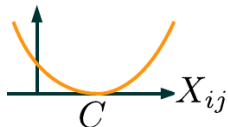
1. The effect of clipping (definition is involved) is small.
2. True matrix M is low-rank.
3. M is “Incoherent” (a small subset of observation is sufficient for estimating the entire matrix)
4. The elements are observed independently with high-enough probability p .

Theorem (Exact recovery for CMC; informal)

With high probability, with a certain algorithm (trace-norm minimization), the true matrix can be recovered exactly.

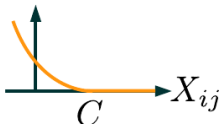
- Ordinary MC [5] : square loss

$$\arg \min_{\mathbf{X}} \frac{1}{2} \sum_{ij : \text{obs}} (\text{Obs}_{ij} - X_{ij})^2 + \mathcal{R}(\mathbf{X})$$



- On the clipped entries, C is wrongly recovered.
- CMC (proposed): square hinge loss

$$\begin{aligned} \arg \min_{\mathbf{X}} \frac{1}{2} \sum_{ij : \text{non-clip. obs.}} (\text{Obs}_{ij} - X_{ij})^2 \\ + \frac{1}{2} \sum_{ij : \text{clipped obs.}} \max(0, \text{Obs}_{ij} - X_{ij})^2 \\ + \mathcal{R}(\mathbf{X}) \end{aligned}$$



Design of regularization term: induce low-rank solution

1. DTr-CMC: Double trace-norm regularization (proposed)
 - ▶ Effect: induce low-rankness in both \mathbf{X} and $\text{Clip}(\mathbf{X})$
 - ▶ Theoretical guarantee is also given (details omitted)
2. Tr-CMC: trace-norm regularization [5]
 - ▶ Effect: induce low-rank solution
3. Fro-CMC: Frobenius norm regularization [4]
 - ▶ Effect: induce low-rank solution

1. DTr-CMC: Double trace-norm regularization

$$\mathcal{R}(\mathbf{X}) = \lambda_1 \|\mathbf{X}\|_{\text{tr}} + \lambda_2 \|\text{Clip}(\mathbf{X})\|_{\text{tr}} \quad \text{Clip} = \min(\cdot, C)$$

- ▶ Effect: induce low-rankness in both \mathbf{X} and $\text{Clip}(\mathbf{X})$
- ▶ Optimization: (approximate) subgradient descent [2]
- ▶ Theoretical guarantee is also given (details omitted)

2. Tr-CMC: trace-norm regularization [5] $\mathcal{R}(\mathbf{X}) := \lambda \|\mathbf{X}\|_{\text{tr}}$

$$\|\mathbf{X}\|_{\text{tr}} = \sum_{l=1}^{\min(n_1, n_2)} \sigma_l \quad (\sigma_l: l\text{-th singular value})$$

- ▶ Effect: induce low-rank solution
- ▶ Optimization: accelerated gradient descent [5]

3. Fro-CMC: Frobenius norm regularization [4]

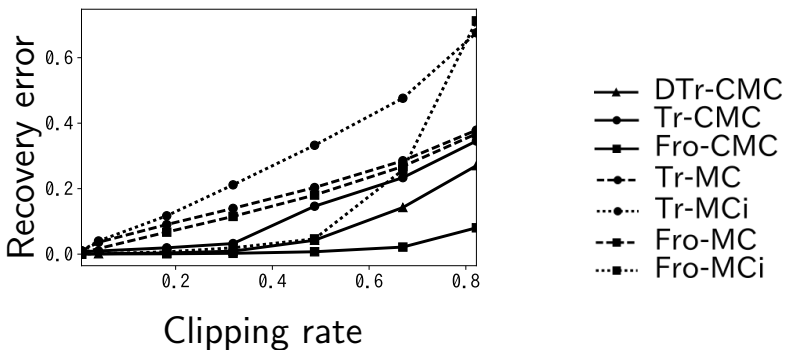
$$\mathcal{R}(\mathbf{P}, \mathbf{Q}) := \lambda_1 \|\mathbf{P}\|_{\text{F}}^2 + \lambda_2 \|\mathbf{Q}\|_{\text{F}}^2 \quad \mathbf{X} = \mathbf{P}\mathbf{Q}^{\text{T}}$$

- ▶ Effect: induce low-rank solution
- ▶ Optimization: (approximate) alternating least squares [4]

1. Experiment with synthetic data
 - ▶ Evaluate the recovery result under a controlled situation (where true value is known)
2. Experiment with real-world data
 - ▶ The true values are unknown \Rightarrow evaluation of recovery is impossible.
(The true values before clipping in real-world data is unknown)
 - ▶ Device for evaluation: evaluate a binary (two-class) classification task to classify entries into “the true value is above threshold or not.”

Experiment 1/3 Experiment with synthetic data

20



- Solid: proposed method, dotted: baseline methods.
- Vary clipping threshold \rightarrow eval. test recovery error.
- Proposed method (solid) is able to estimate the true matrix with small error of order 10^{-2} even when there is 70% clipping.

Experiment 2/3 Experiment with real-world data (1)

f_1 value	DTr-CMC	Fro-CMC	Fro-MC	Tr-CMC	Tr-MC	(baseline)
Film Trust	0.47 (0.01)	0.35 (0.01)	0.27 (0.01)	0.36 (0.00)	0.22 (0.00)	0.41 (0.00)
Movielens 100K	0.39 (0.00)	0.41 (0.00)	0.21 (0.01)	0.40 (0.00)	0.12 (0.00)	0.35 (0.00)

- Learning after artificially clipping real data (★ 5 → ★ 4)
- Classify true ratings into “★ 5 \geq ” and “ \leq ★ 4”
- Proposed method estimates **the true value better**
- (Baseline: a classifier which unconditionally outputs +1)

Experiment 3/3 Experiment with real-world data (2)

f_1 value	DTr-CMC	Fro-CMC	Fro-MC	Tr-CMC	Tr-MC	(baseline)
Film Trust	0.46 (0.01)	0.40 (0.01)	0.35 (0.01)	0.39 (0.00)	0.35 (0.01)	0.41 (0.00)
Movielens 100K	0.38 (0.00)	0.41 (0.01)	0.38 (0.01)	0.40 (0.00)	0.38 (0.00)	0.35 (0.00)

- Learning from real data (without artificial clipping) (★ 1~★ 5)
- Classify true ratings into “★ 5 \geq ” and “ \leq ★ 4”
- **Robustness to ceiling effect** improves the detection power of high-rating entries.
- (Baseline: a classifier which unconditionally outputs +1)

- Problem setting? → Recover matrix from ceiling effects
- Recovery possible? → **Exact recovery possible** under conditions
- How to recover? → Minimize **square hinge loss** + regularization term
- Experimentally? → Resilience to ceiling effect may improve **recommendation systems!**

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

True values
(unknown)

Obs.



	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Observed data
(clip @ 10)

Rec.



4.0	7.0	4.0	7.0	4.0
-0.0	3.0	6.0	14.9	11.9
4.0	6.0	2.0	2.0	0.0
2.0	6.0	7.0	15.9	11.9
8.0	13.0	6.0	9.0	4.0

After
completion



- Let $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top$ (Singular value decomposition).
- Define *coherence* by $\mu_0 := \max \left\{ \frac{n_1}{r} \mu^U(\mathbf{M}), \frac{n_2}{r} \mu^V(\mathbf{M}) \right\}$
 - ▶ Here, $\mu^U(\mathbf{M}) := \max_{i \in [n_1]} \|\mathbf{U}_{i,\cdot}\|^2$,
 $\mu^V(\mathbf{M}) := \max_{j \in [n_2]} \|\mathbf{V}_{j,\cdot}\|^2$, $r = \text{rank}(\mathbf{M})$.
- Joint coherence is defined by $\mu_1 := \sqrt{\frac{n_1 n_2}{r}} \|\mathbf{U}\mathbf{V}^\top\|_\infty$
- \mathbf{M} is *incoherent* if both μ_0 and μ_1 are small.

- $\mathcal{B} := \{(i, j) : M_{ij} < C\}$

-

$$T := \text{span}(\{\mathbf{u}_k \mathbf{y}^\top : k \in [r], \mathbf{y} \in \mathbb{R}^{n_2}\} \cup \{\mathbf{x} \mathbf{v}_k^\top : k \in [r], \mathbf{x} \in \mathbb{R}^{n_1}\})$$

- $(\mathcal{P}^*(\mathbf{Z}))_{ij} := \mathbf{1}\{M_{ij} < C\} Z_{ij} + \mathbf{1}\{M_{ij} = C\} (Z_{ij})_+$

- $\rho_F := \sup_{\mathbf{Z} \in T \setminus \{\mathbf{O}\}: \|\mathbf{Z}\|_F \leq \|\mathbf{U}\mathbf{V}^\top\|_F} \frac{\|\mathcal{P}_T \mathcal{P}^*(\mathbf{Z}) - \mathbf{Z}\|_F}{\|\mathbf{Z}\|_F}$

- $\rho_\infty := \sup_{\mathbf{Z} \in T \setminus \{\mathbf{O}\}: \|\mathbf{Z}\|_\infty \leq \|\mathbf{U}\mathbf{V}^\top\|_\infty} \frac{\|\mathcal{P}_T \mathcal{P}^*(\mathbf{Z}) - \mathbf{Z}\|_\infty}{\|\mathbf{Z}\|_\infty}$

- $\rho_{\text{op}} := \sqrt{r} \mu_1 \left(\sup_{\|\mathbf{Z}\|_{\text{op}} \leq \sqrt{n_1 n_2} \|\mathbf{U}\mathbf{V}^\top\|_{\text{op}}} \frac{\|\mathcal{P}^*(\mathbf{Z}) - \mathbf{Z}\|_{\text{op}}}{\|\mathbf{Z}\|_{\text{op}}} \right)$

- $\nu_{\mathcal{B}} := \|\mathcal{P}_T \mathcal{P}_{\mathcal{B}} \mathcal{P}_T - \mathcal{P}_T\|_{\text{op}}$

Can we account for floor effect in addition to ceiling effect?

- Yes, the result is extendable to floor effect (cf. the paper).
- Element-wise thresholds are also allowed.

How to determine if recovery is possible 28

- The assumptions can be checked only after seeing the true matrix.
- However, there are some intuitions what kind of matrix is possible to recover.
 - ▶ Low-rank = consists of a small number of components
 - ▶ A few (column-/row-wise) common factors almost determine the matrix entries.
 - ▶ In other words, there are similar rows/columns.
 - ▶ The space spanned by the singular vectors of \mathbf{M} is not aligned with the indicator-matrices of the indices which are to be clipped.
 - Roughly speaking, rank-one matrices of SVD (the components of \mathbf{M}) have support all over the indices (if there are sparse components which may have large values on clipped indices, the recovery is impossible).

- No theoretically justified method for hyper-parameter selection in recovery problems.
 - ▶ In synthetic data experiment, we selected the parameter with the smallest difference between the data and the clipped version of the estimated matrix.
- In the real data experiment, the final performance can be computed. Therefore, we used the one with the best performance on a held-out validation indices.
 - ▶ Similarly, in recommendation systems, the final performance measure is likely available for hyper-parameter selection.

Future work for clipped matrix completion

- Characterize necessary condition for recovery.
- Develop algorithms to perform artificial clipping to disable a recovery by arbitrary method.

Trace-norm minimization is the algorithm defined as below.

$$\arg \min_{\mathbf{X}} \|\mathbf{X}\|_{\text{tr}} \text{ s.t. } \begin{cases} \mathcal{P}_{\Omega \setminus \mathcal{C}}(\mathbf{X}) = \mathcal{P}_{\Omega \setminus \mathcal{C}}(\mathbf{M}_{\Omega}^c), \\ \mathcal{P}_{\mathcal{C}}(\mathbf{M}_{\Omega}^c) \leq \mathcal{P}_{\mathcal{C}}(\mathbf{X}), \end{cases}$$

Here, $\Omega := \{(i, j) : \text{observed}\}$ and $\mathcal{C} := \{(i, j) \in \Omega : M_{ij}^c = C\}$.

- precision : Among those predicted “yes,” the fraction of true “yes.” (ratio of precise predictions)
- recall : Among those with true “yes,” the fraction of those predicted “yes.” (ratio of correctly recalled true “yes”)
- $f_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$
- f_1 was used because in this binary classification, the challenge is to **extrapolate** to a large value from observed small values. Therefore, recall is considered as the difficult part.

- [1] Encyclopedia of research design.
- [2] Haim Avron, Satyen Kale, Shiva Prasad Kasiviswanathan, and Vikas Sindhwani.
Efficient and practical stochastic subgradient descent for nuclear norm regularization.
In *Proceedings of the 29th International Conference on Machine Learning*, pages 1231–1238.
- [3] Emmanuel J. Cands and Yaniv Plan.
Matrix completion with noise.
98(6):925–936.
- [4] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi.
Low-rank matrix completion using alternating minimization.
In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pages 665–674.
- [5] Kim-Chuan Toh and Sangwoon Yun.
An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems.
6:615–640.