Matrix Completion

Yuepeng Yang

University of Chicago, Department of Statistics

Oct 12, 2024

Table of Contents

Netflix Prize

- In 2006, Netflix offers \$1 million to improve 10% over its own algorithm Cinematch
- Data consists of 100,480,507 ratings, 480,189 users, 17,770 movies

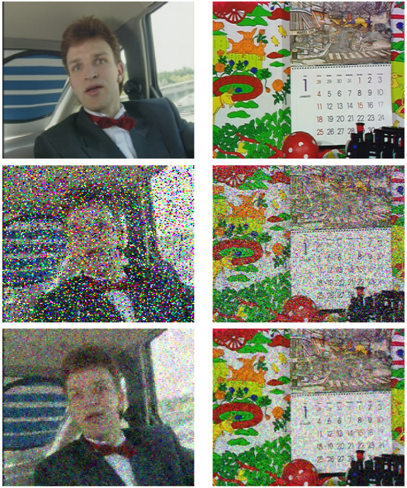
	Movie A	Movie B	Movie C	Movie D	Movie E
Alice	?	4	1	?	?
Bob	?	2	?	8	1
Charlie	?	?	?	?	?
David	9	2	?	?	1
Euler	?	2	?	5	?

- Latent structure allow us to infer the unobserved entries

Drug effect prediction

Target	Drug A	Drug B	Drug C	Drug D	Drug E
Protein A	?	✓	?	×	?
Protein B	?	×	?	✓	?
Protein C	?	?	×	?	?
Protein D	✓	×	×	?	✓
Protein E	?	×	?	✓	?

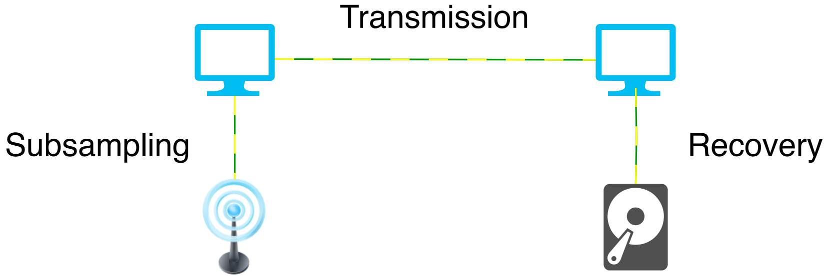
Image / Video restoring



(a) Sequence 1 (b) Sequence 2

figure credit: Ji et al., 2010. Robust video denoising using Low rank matrix completion.

Signal compressing and transmission



The matrix completion problem

Observe

$$\boldsymbol{M} = \mathcal{P}_{\Omega}(\boldsymbol{M}^{\star})$$

Goal: estimate $\boldsymbol{M}^{\star}$ given $\boldsymbol{M}$ and Ω.

$$\begin{bmatrix} \checkmark & ? & ? & ? & \checkmark & ? \\ ? & ? & \checkmark & \checkmark & ? & ? \\ \checkmark & ? & ? & \checkmark & ? & ? \\ ? & ? & \checkmark & ? & ? & \checkmark \\ \checkmark & ? & ? & ? & ? & ? \\ ? & \checkmark & ? & ? & \checkmark & ? \\ ? & ? & \checkmark & \checkmark & ? & ? \end{bmatrix}$$

The observation set Ω

Identifiability

Observe:

	Movie A	Movie B
Alice	?	4
Bob	1	2

Solution 1:

	Movie A	Movie B
Alice	2	4
Bob	1	2

Solution 2:

	Movie A	Movie B
Alice	10	4
Bob	1	2

Identifiability

Observe:

	Movie A	Movie B
Alice	?	4
Bob	1	2

Solution 1:

	Movie A	Movie B
Alice	2	4
Bob	1	2

Solution 2:

	Movie A	Movie B
Alice	10	4
Bob	1	2

Which solution is the correct one?

Singular value decomposition

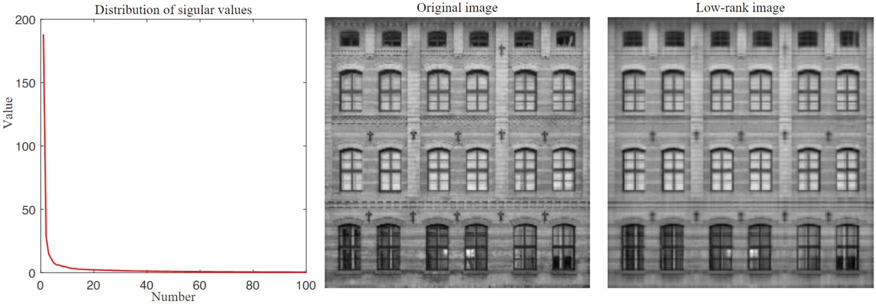
Singular value decomposition:

$$\boldsymbol{M} = \sum_{i=1}^{n} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \boldsymbol{U}\Sigma\boldsymbol{V}^\top$$

where $\boldsymbol{u}_i, \boldsymbol{v}_i$ are orthonormal basis

Intuitively, it iteratively selects the rank-1 matrix the represents the most signal in $\boldsymbol{M}$ and σ_i is the signal strength

Low-rank assumption



comparison between the image and a rank-10 approximation
figure credit: Li et al., 2019. A Survey on Matrix Completion: Perspective of Signal Processing.

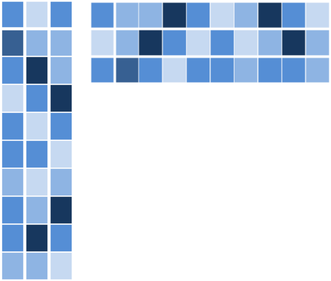
The low-rank matrix completion problem

Observe $n_1 \times n_2$ matrix

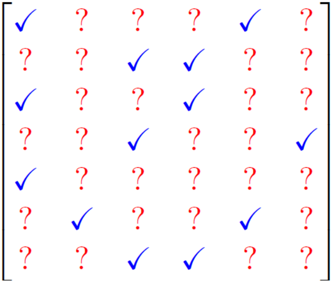
$$\boldsymbol{M} = \mathcal{P}_\Omega(\boldsymbol{M}^\star)$$

where $\boldsymbol{M}^\star$ is a rank-r matrix, i.e., it can be written as $\boldsymbol{M}^\star = \boldsymbol{X}^\star \boldsymbol{Y}^{\star\top}$ for matrix $\boldsymbol{X}^\star \in \mathbb{R}^{n_1 \times r}$ and matrix $\boldsymbol{Y}^\star \in \mathbb{R}^{n_2 \times r}$.

Goal: estimate $\boldsymbol{M}^\star$ given $\boldsymbol{M}$ and Ω.



The decomposition of the hidden rank-r matrix $\boldsymbol{M}^\star$



The observation set Ω

Netflix Prize with a low-rank representation

	HF 1	HF 2
Alice	5	4
Bob	1	2
Charlie	5	1
David	9	2
Euler	7	2

Movies	A	B	C	D	E
HF 1	1	4	3	6	4
HF 2	5	2	5	8	1

HF = Hidden Feature

Table of Contents

The task

Find a matrix $\boldsymbol{L} \in \mathbb{R}^{n_1 \times n_2}$ such that

$$\boldsymbol{M} = \mathcal{P}_\Omega(\boldsymbol{L})$$

and

$$\text{rank}(\boldsymbol{L}) = r$$

We also assume each entry is observed with i.i.d. probability p.

Rank minimization

$$\min_{\boldsymbol{L} \in \mathbb{R}^{n_1 \times n_2}} \ \operatorname{rank}(\boldsymbol{L})$$

such that

$$\boldsymbol{M} = \mathcal{P}_{\Omega}(\boldsymbol{L})$$

Rank minimization

$$\min_{\boldsymbol{L} \in \mathbb{R}^{n_1 \times n_2}} \quad \text{rank}(\boldsymbol{L})$$

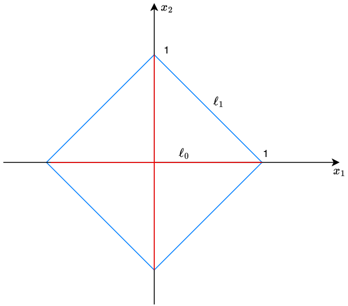
such that

$$\boldsymbol{M} = \mathcal{P}_{\Omega}(\boldsymbol{L})$$

rank minimization is NP-hard!

Relaxing the rank

- Rank as a function is hard to optimize
- $\text{rank}(\boldsymbol{L}) = \|\sigma(\boldsymbol{L})\|_0$: number of non-zero singular values
- $\|\boldsymbol{L}\|_\star = \|\sigma(\boldsymbol{L})\|_1$: sum of singular values



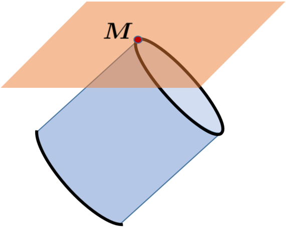
L1 and L0 unit balls

Nuclear norm minimization

$$\min_{\boldsymbol{L}\in\mathbb{R}^{n_1\times n_2}} \|\boldsymbol{L}\|_* \quad \text{s.t.} \quad \boldsymbol{M} = \mathcal{P}_\Omega(\boldsymbol{L})$$

It is a convex program and polynomial-time computable

The minimizer $\widehat{\boldsymbol{L}}$ recovers $\boldsymbol{M}^\star$ exactly under reasonable conditions



Intersection of $\{\boldsymbol{L} : \boldsymbol{M} = \mathcal{P}_\Omega(\boldsymbol{L})\}$ and level set of $\|\boldsymbol{L}\|_*$

figure credit: Yuejie Chi

Another reformulation

A rank-r matrix $\boldsymbol{L}$ can be written as $\boldsymbol{L} = \boldsymbol{X}\boldsymbol{Y}^\top$ for some $n_1 \times r$ matrix $\boldsymbol{X}$ and $n_2 \times r$ matrix $\boldsymbol{Y}$.

$$\boldsymbol{M} = \mathcal{P}_\Omega(\boldsymbol{L}) \qquad \text{s.t.} \qquad \operatorname{rank}(\boldsymbol{L}) = r$$

Another reformulation

A rank-r matrix $\boldsymbol{L}$ can be written as $\boldsymbol{L} = \boldsymbol{X}\boldsymbol{Y}^\top$ for some $n_1 \times r$ matrix $\boldsymbol{X}$ and $n_2 \times r$ matrix $\boldsymbol{Y}$.

$$\min \ \|\boldsymbol{M} - \mathcal{P}_\Omega(\boldsymbol{X}\boldsymbol{Y}^\top)\|_\mathsf{F}^2$$

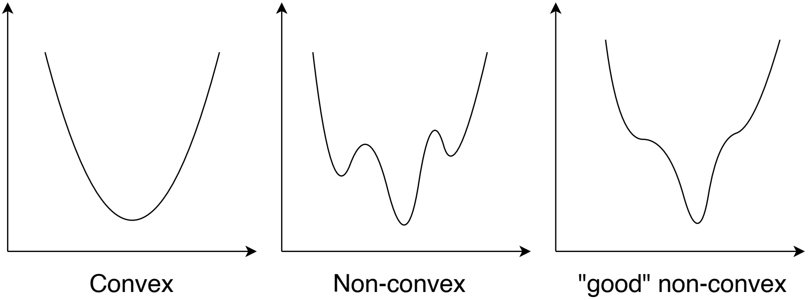
Another reformulation

A rank-r matrix $\boldsymbol{L}$ can be written as $\boldsymbol{L} = \boldsymbol{X}\boldsymbol{Y}^\top$ for some $n_1 \times r$ matrix $\boldsymbol{X}$ and $n_2 \times r$ matrix $\boldsymbol{Y}$.

$$\min \ \|\boldsymbol{M} - \mathcal{P}_\Omega(\boldsymbol{X}\boldsymbol{Y}^\top)\|_\mathsf{F}^2$$

A non-convex program

The two faces of non-convexity



Ge, Lee, and Ma, 2016. Neurips.

Matrix Completion has No Spurious Local Minimum.

i.e. all local minima is global

Algorithm: Gradient descent with spectral initialization

- Spectral initialization to get a good initial guess
  - Intuition:

$$\mathbb{E}\left[\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{M}^\star)\right] = \boldsymbol{M}^\star = \boldsymbol{X}^\star \boldsymbol{Y}^{\star\top}$$

  - Top-r SVD on $\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{L}^\star)$

$$\frac{1}{p}\mathcal{P}_\Omega(\boldsymbol{M}^\star) = \boldsymbol{U}_0 \boldsymbol{\Sigma}_0 \boldsymbol{V}_0^\top = \boldsymbol{U}_0 \boldsymbol{\Sigma}_0^{1/2} \boldsymbol{\Sigma}_0^{1/2} \boldsymbol{V}_0^\top = \boldsymbol{X}_0 \boldsymbol{Y}_0^\top$$

Algorithm: Gradient descent with spectral initialization

- Gradient descent

$$F(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{2p} \|\boldsymbol{M} - \mathcal{P}_{\Omega}(\boldsymbol{X}\boldsymbol{Y}^{\top})\|_{\mathrm{F}}^2 + \frac{1}{8} \|\boldsymbol{X}^{\top}\boldsymbol{X} - \boldsymbol{Y}^{\top}\boldsymbol{Y}\|_{\mathrm{F}}^2$$

$$\begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{Y}_{t-1} \end{bmatrix} - \eta_t \begin{bmatrix} \nabla_{\boldsymbol{X}} F(\boldsymbol{X}_{t-1}, \boldsymbol{Y}_{t-1}) \\ \nabla_{\boldsymbol{Y}} F(\boldsymbol{X}_{t-1}, \boldsymbol{Y}_{t-1}) \end{bmatrix}$$

- $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = \sum_{i,j} A_{ij}^2$
- Solvable in linear time and converges to ground truth under reasonable conditions

Comparison

Nuclear norm minimization	GD with spectral init
convex	non-convex
don't need to know r	need to know r
polynomial time	linear time
exact recovery	exact recovery

Table of Contents

Missing patterns

Cai, Cai, and Zhang, 2016. Structured Matrix Completion with Applications to Genomic Data Integration. JASA.
Choi and Yuan, 2024. Matrix Completion When Missing Is Not at Random and Its Applications in Causal Panel Data Models. JASA.

	A	B	C	D
Alice	?	✓	?	?
Bob	?	✓	?	✓
Charlie	?	?	✓	?
David	✓	?	?	✓

Random observation

	A	B	C	D
Alice	✓	✓	✓	✓
Bob	✓	✓	✓	✓
Charlie	✓	✓	?	?
David	✓	✓	?	?

Structured observation

Models on observed data

Chen et al. 2023. Statistical inference for noisy incomplete binary matrix. JMLR.

	A	B	C	D
Alice	0.1	0.64	0.87	0.16
Bob	0.54	0.34	0.56	0.23
Charlie	0.23	0.52	0.53	0.82
David	0.7	0.02	0.94	0.33

Latent probability

	A	B	C	D
Alice	?	✓	?	?
Bob	?	×	?	×
Charlie	×	?	✓	?
David	✓	?	?	✓

Binary observation

Noisy data

$$\boldsymbol{M} = \mathcal{P}_{\Omega}(\boldsymbol{L}^{\star} + \boldsymbol{E})$$

- $\boldsymbol{E}$ is entrywise independent Gaussian noise
- $\boldsymbol{E}$ is entrywise independent sub-exponential noise
  - Farias, Li, and Peng, 2022. Uncertainty Quantification For Low-Rank Matrix Completion With Heterogeneous and Sub-Exponential Noise. AISTATS
- $\boldsymbol{E}$ is entrywise independent heavy-tailed noise.
  - Wang and Fan, 2023. Robust Matrix Completion with Heavy-tailed Noise. JASA.

More structure in the data

Non-negative matrix factorization:

$$\boldsymbol{L}^\star = \boldsymbol{X}^\star \boldsymbol{Y}^{\star\top}$$

$\boldsymbol{L}^\star, \boldsymbol{X}^\star, \boldsymbol{Y}^\star$ all positive.

	Movie A	Movie B	Movie C	Movie D	Movie E
Alice	?	4	1	?	?
Bob	?	2	?	8	1
Charlie	?	?	?	?	?
David	9	2	?	?	1
Euler	?	2	?	5	?

Covariate Information

$$\boldsymbol{M} = \mathcal{P}_{\Omega}(\boldsymbol{L}^{\star} + \boldsymbol{E})$$

$$\boldsymbol{L}^{\star} = \boldsymbol{X}\beta + \boldsymbol{B}$$

Mao, Chen, and Wong, 2019. Matrix completion with covariate information. JASA.

	A	B	C	D
Alice	?	4	1	?
Bob	?	2	?	8
Charlie	?	?	?	?
David	9	2	?	?

	A	B	C	D
Genre	Action	Sci-fi	Fantasy	Horror
Year	1995	2020	2024	2016
Budget	100k	20m	200m	1k

Algorithms with desired properties

- We want our algorithm to be fast, low memory cost, parallelizable, free of tuning parameters, etc.
- Recht and Ré, 2013. Parallel stochastic gradient algorithms for large-scale matrix completion. MPC.
- Yang and Ma, 2023. Optimal tuning-free convex relaxation for noisy matrix completion. IEEE TIT.

Uncertainty quantification

- How variable the estimate is?
- Chen et al., 2019. Inference and uncertainty quantification for noisy matrix completion. PNAS.

	A	B	C	D
Alice	(1.4, 2.6)	4	1	(1.1, 1.3)
Bob	(2.5, 4.5)	2	(9.1, 10)	8
Charlie	(7.8, 9.1)	(2.1, 5.7)	6	5
David	9	2	(5.2, 5.5)	(2.6, 4.1)

95% CI estimate

Table of Contents

Summary

- Applications
  - recommendation system, data imputation, image recovery...
- Problem formulation of matrix completion
- Solution 1: nuclear norm minimization
  - convex relaxation
- Solution 2: gradient descent with spectral initialization
  - Benign landscape of non-convex optimization
- Snippet of active research questions