Introduction to Matrix Completion

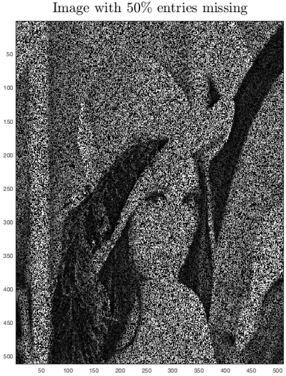
Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: September 21, 2019
Last update : February 5, 2020

- **Incomplete matrix** : given a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ that not all the values in $\mathbf{M}$ are observed.



- The goal of Matrix Completion (MC) : recover those missing values.

Why the data is incomplete?

The incompleteness of data comes from various sources

- Caused by nature
  - Hardware failure (e.g. sensors)
  - Blocked by obstacle : in Earth imaging, the cloud blocks the view of the satellite and thus creating a large area with white in colour. By viewing the region blocked by cloud as non-data, we have an incomplete image.
- Caused by human
  - The Netflix problem / use-rating data : most users only rate a few movies but not every movie they watched
  - Censorship due to political reasons

The problem setting of MC

In a MC problem, we are given :

- A partially observed matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$
- An index set Ω labelling the observed entries where

$$(i, j) \in \Omega \iff (i, j) \text{ entry is observed.}$$

  where $|\Omega| \leq mn$ is the number of observed entries in $\mathbf{M}$.
- Based on Ω, we have the complement set Ω^c that

$$(i, j) \in \Omega^c \iff (i, j) \text{ entry is not observed / is missing.}$$

- We construct an estimator $\mathbf{X}$ of $\mathbf{M}$ such that, for each location (i, j):
  - If $(i, j) \in \Omega$, we want $\mathbf{X}(i, j) = \mathbf{M}(i, j)$
  - If $(i, j) \in \Omega^C$, we want to estimate (impute) the value $\mathbf{X}(i, j)$ such that this estimation "makes sense"
- However, what does this "makes sense" mean??

Casting MC as an optimization problem

A criteria that "makes sense" is low rank : you want the estimation to be the one with the lowest complexity out of all guesses (Occam's razor).

$$\underset{\mathbf{X}}{\operatorname{argmin}} \operatorname{rank}(\mathbf{X}) \text{ s.t. } \mathbf{X}(i,j) = \mathbf{M}(i,j), \ \forall (i,j) \in \Omega$$

- The problem is a equality-constrained optimization problem
- red part : among all possible $\mathbf{X}$, find the one that has the lowest rank
- blue part : the entries of $\mathbf{X}$ for $(i,j) \in \Omega$ has to be consistent to those in $\mathbf{M}$
- The constraint $\mathbf{X}(i,j) = \mathbf{M}(i,j), \ \forall (i,j) \in \Omega$ can also be compactly denoted as $\mathbf{X}_\Omega = \mathbf{M}_\Omega$

That is, we want to find an $\mathbf{X}$, such that it is as low rank as possible, subject to the constraint that entries in $\mathbf{X}$ agree with the observed ones in $\mathbf{M}$.

NP-hardness of the rank minimization problem

The problem

$$\underset{\mathbf{X}}{\operatorname{argmin}} \operatorname{rank}(\mathbf{X}) \text{ s.t. } \mathbf{X}_\Omega = \mathbf{M}_\Omega$$

is NP-hard : as $\operatorname{rank}(\mathbf{X})$ is the l_0 norm on singular values of $\mathbf{X}$

$$\operatorname{rank}(\mathbf{X}) = \|\operatorname{diag}(\Sigma)\|_0 = \text{ number of non-zero singular value of } \mathbf{X}.$$

As l_0-norm problem has combinatorial complexity, so this problem is NP-Hard.

Under some technical assumptions, the problem above can be solved by solving an relaxed problem using the *Nuclear norm*.

Nuclear norm

- The nuclear norm of $\mathbf{X}$ is the sum of singular value of $\mathbf{X}$.

$$\underbrace{\|\mathbf{X}\|_* := \sum_i |\sigma_i|}_{\text{definition of nuclear norm}} = \sum_i \sigma_i$$

  in which the absolute sign can be dropped as singular values are all non-negative

- It can be shown that, nuclear norm is the tightest convex relaxation of the rank function within the unit norm ball. See the proof here.

Nuclear Norm minimization problem

Using the nuclear norm, we have

$$\underset{\mathbf{X}}{\operatorname{argmin}} \ \|\mathbf{X}\|_* \ \text{s.t.} \ \mathbf{X}_\Omega = \mathbf{M}_\Omega$$

- Under some technical assumptions, the solution of this problem is the same as the solution of the rank minimization problem, so solving this problem is meaningful
- As nuclear norm is convex, this problem is much easier to solve than the NP-hard rank minimization problem
- This problem can be solved by various approaches
  - Majorization-minimization method
  - Proximal point method
  - Augmented Lagrangian method
  - Interior point method
  - Semi-definite programming method

Variations on the problem setting : noisy completion

- The constraint $\mathbf{X}_\Omega = \mathbf{M}_\Omega$ basically means that the solution has to "hard-code" all entries in Ω as $\mathbf{M}_\Omega$, which is OK for noiseless data
- If data is (highly) noisy, hard-coding is harmful : you learn the noise
- This suggest the use of soft penalty

$$\underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\|_* + \frac{\lambda}{2} \sum_{(i,j)\in\Omega} \Big(\mathbf{X}(i,j) - \mathbf{M}(i,j)\Big)^2$$

  where $\lambda > 0$ is a parameter. The model means minimize the nuclear norm of $\mathbf{X}$ such that all $\mathbf{X}_\Omega$ is not too far away from $\mathbf{M}_\Omega$.
- Compact notation of the above is

$$\underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\|_* + \frac{\lambda}{2} \|\mathbf{X}_\Omega - \mathbf{M}_\Omega\|_F^2$$

Variations on the problem setting : more general case

The more general case of the MC problem reads

$$\underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\|_* \text{ s.t. } \mathcal{A}(\mathbf{X}) = \mathbf{b}$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is a generic linear operator and $\mathbf{b} \in \mathbb{R}^p$ where $p < mn$.

In fact $X_\Omega = \mathbf{M}_\Omega$ is just a special case of $\mathcal{A}(\mathbf{X}) = \mathbf{b}$: here $p = |\Omega|$, $\mathbf{b} = \text{vec}(\mathbf{M}_\Omega)$ and $\mathcal{A}$ is a operator consists of vectorization based on the structure of Ω.

Recoverability of MC problems

- Not all MC problems are solvable : for example, if only 1 pixel is observed, then it is almost impossible to recover the original $\mathbf{M}$, unless the true $\mathbf{M}$ is a constant matrix (all entries share the same value as the observed one).
- What does "solvable" means : we found the "right thing" – assume there is a ground truth, and the entries of $\mathbf{X}_{\Omega^C}$ are exactly the ground truth (or very close to them).
- If the entries of $\mathbf{X}_{\Omega^C}$ are exactly (or very close to) the ground truth, we said the $\mathbf{X}$ recovers the missing values correctly.
- There is a fundamental limit on the number $|\Omega|$ such that the problem is solvable, or it is recoverable for those entries of $\mathbf{M}_{\Omega}$.
- We will discuss the recoverability issue in other documents.

Last page - summary

What we discussed : basic understanding of matrix completion

- Problem setting
- Problem formulation

Not discussed – topics in matrix completion

- Recoverability of MC problem – how many samples are need to recover the ground truth
- How to actually solve the MC minimization problem – algorithm design

End of document