Solving Nuclear Norm Minimization by Majorization-Minimization

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

The MC problem

The noiseless nuclear norm based MC problem

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

- $\mathbf{X} \in \mathbb{R}^{m \times n}$ is the optimization variable
- $\Omega \subset [1,\ldots,m] \times [1,\ldots,n]$ is an index set storing the location of the observed entries, it is given
- $\mathbf{M} \in \mathbb{R}^{m \times n}$ is the observed data, it is given

The noisy nuclear norm based MC problem

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

where $\lambda > 0$ is a parameter related to the noise level.

Majorizing the penalty part

Consider the noiseless case[1]. We now derive the MM formulation for the MC problem. See here for background of MM.

Let $f(\mathbf{X}) = \|\mathbf{X}\|_*$. The following is always true

$$f(\mathbf{X}) \leq g(\mathbf{X};\mathbf{Y}) := f(\mathbf{X}) + \underbrace{\|\mathbf{X}-\mathbf{Y}\|_F^2}_{\text{always } \geq 0}$$

Equality between f, g holds when $\mathbf{Y} = \mathbf{X}$. Thus $g(\mathbf{X};\mathbf{Y})$ is a majorizer of $f(\mathbf{X})$.
In this sense we have the following new problem

$$\underset{\mathbf{X}}{\text{argmin}}\ g(\mathbf{X};\mathbf{Y}) = \|\mathbf{X}\|_* + \frac{1}{2\lambda}\|\mathbf{X}-\mathbf{Y}\|_F^2 \text{ s.t. } \mathbf{X}(i,j) = \mathbf{M}(i,j)\ \forall (i,j) \in \Omega$$

where $\lambda > 0$ is a constant.

[1] For the noisy case, just add the quadratic term. The idea on how to solve the problem remains the same.

The purpose of the auxiliary variable $\mathbf{Y}$

By introducing $\mathbf{Y}$, the constraint $\mathbf{X}_\Omega = \mathbf{M}_\Omega$ can be decoupled from the variable $\mathbf{X}$. As

$$\mathbf{X} = \operatorname*{argmin}_{\mathbf{X}} \|\mathbf{X}\|_* + \frac{1}{2\lambda}\|\mathbf{X} - \mathbf{Y}\|_F^2$$

has no constraint on $\mathbf{X}$, we can use whatever method for unconstrained problem to solve this.

On the other hand, for the definition of $\mathbf{Y}$, we can let it be

$$\mathbf{Y}(i,j) = \begin{cases} \mathbf{M}(i,j) & (i,j) \in \Omega \\ \mathbf{X}(i,j) & (i,j) \in \Omega^c \end{cases}$$

i.e. for observed entries, we put $\mathbf{Y}(i,j)$ as the corresponding values in $\mathbf{M}$. For unobserved entries, we put $\mathbf{Y}(i,j)$ as the corresponding values in $\mathbf{X}$.

The whole MM algorithm

Algorithm 1: MM algorithm for noiseless MC

Input $\mathbf{M} \in \mathbb{R}^{m \times n}$, Ω, parameter $\lambda > 0$;

Initialize $\mathbf{X}_0 \in \mathbb{R}^{m \times n}$;

for $k = 1, 2, \ldots$ **do**

1. Compute $\mathbf{Y}_k$ from $(\mathbf{M}, \mathbf{X}_{k-1})$ as

$$\mathbf{Y}_k(i,j) = \begin{cases} \mathbf{M}(i,j) & (i,j) \in \Omega \\ \mathbf{X}_{k-1}(i,j) & (i,j) \in \Omega^c \end{cases}$$

2. Update $\mathbf{X}_k$ as follows

$$\mathbf{X}_{k+1} = \underset{\mathbf{X}}{\operatorname{argmin}} \ \|\mathbf{X}\|_* + \frac{1}{2\lambda}\|\mathbf{X} - \mathbf{Y}_k\|_F^2$$

end

What is left is how to solve the sub-problem on $\mathbf{X}$.

Solving the sub-problem

The sub-problem

$$\underset{\mathbf{X}}{\operatorname{argmin}} \left(\|\mathbf{X}\|_* + \frac{1}{2\lambda}\|\mathbf{X} - \mathbf{Y}_k\|_F^2 \right)$$

is equivalent to

$$\underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{\lambda} \left(\lambda\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}_k\|_F^2 \right)$$

As $\underset{\mathbf{X}}{\operatorname{argmin}}$ is independent of scaling, we can drop the $\frac{1}{\lambda}$ and get

$$\underset{\mathbf{X}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}_k\|_F^2 + \lambda\|\mathbf{X}\|_*$$

which has close form solution called *singular value thresholding operator* (SVT).

The final algorithm

Algorithm 2: MM algorithm for MC

Input $\mathbf{M} \in \mathbb{R}^{m \times n}$, Ω, parameter $\lambda > 0$;

Initialize $\mathbf{X}_0 \in \mathbb{R}^{m \times n}$;

for $k = 1, 2, \ldots$ **do**

1. Compute $\mathbf{Y}_k$ from $(\mathbf{M}, \mathbf{X}_{k-1})$ as

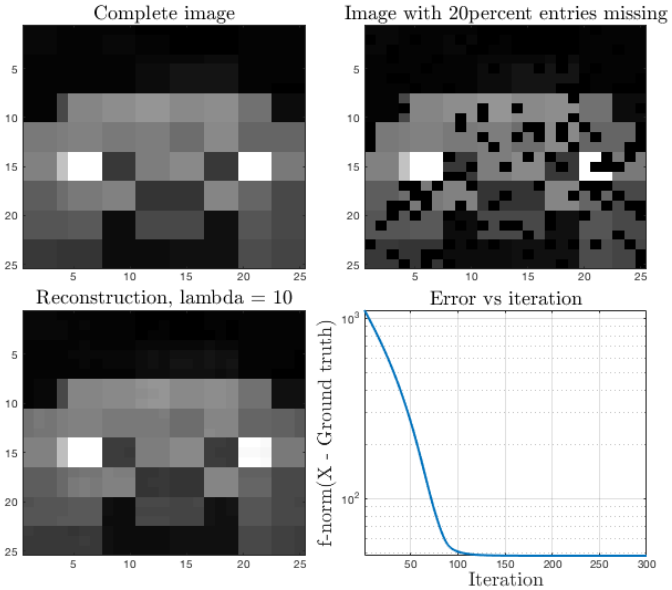
$$\mathbf{Y}_k(i,j) = \begin{cases} \mathbf{M}(i,j) & (i,j) \in \Omega \\ \mathbf{X}_{k-1}(i,j) & (i,j) \in \Omega^c \end{cases}$$

2. Perform SVD on $\mathbf{Y}$: $\mathbf{U}\Sigma\mathbf{V}^\top = \text{SVD}(\mathbf{Y})$.
3. $\mathbf{X}_{k+1} = \mathbf{U}[\Sigma - \lambda\mathbf{I}]_+\mathbf{V}^\top$

end

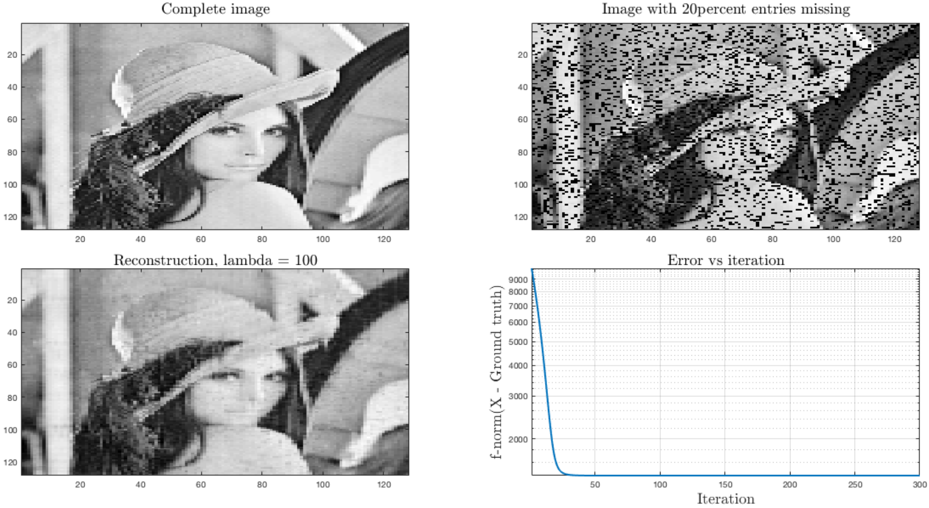
Simple MATLAB code on two toy examples

On simple problem



Note : the regularization parameter is not tuned in this example.

On more complicated problem



Note : the regularization parameter is not tuned in this example.

End of document