MA3113: Topics in Mathematical Image Processing I
Principal Component Pursuit

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

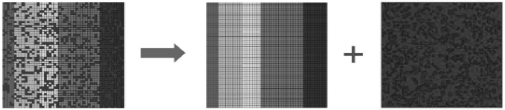
First version: September 30, 2022/Last updated: March 18, 2026

Sparse plus low rank matrix decomposition

Let $M \in \mathbb{R}^{m\times n}$ be a given grayscale image. Suppose that M is the superposition of a low-rank component L and a sparse component S,

$$M = L + S.$$

We are interested in finding the low-rank image L, which has high repeatability along horizontal or vertical directions.



(schematic diagram)

The sparse plus low rank decomposition problem can be formulated as the constrained minimization problem:

$$\min_{L,S}\left(\text{rank}(L) + \lambda\|S\|_0\right) \quad \text{subject to} \quad M = L + S,$$

where $\lambda > 0$ is a tuning parameter and $\|S\|_0$ denotes the number of non-zero entries in S. *The problem is not convex.*

Note: Let $f(A) := \text{rank}(A)$. Then f is neither convex nor concave.

The principal component pursuit problem

We approximate the sparse plus low rank decomposition problem by the following *principal component pursuit (PCP) problem:*

$$\min_{\boldsymbol{L},\boldsymbol{S}} \left(\|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \right) \quad \text{subject to} \quad \boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S},$$

where $\|\boldsymbol{L}\|_*$ is the nuclear (Ky Fan/樊"土畿") norm of $\boldsymbol{L}$ defined as

$$\|\boldsymbol{L}\|_* := \sum_{i=1}^{r} \sigma_i,$$

and $r \in \mathbb{N}^+$ is the rank of $\boldsymbol{L}$ and σ_i are the singular values of $\boldsymbol{L}$, and $\|\boldsymbol{S}\|_1$ denotes the ℓ^1-norm of $\boldsymbol{S}$ *(seen as a long vector in $\mathbb{R}^{mn}$),*

$$\|\boldsymbol{S}\|_1 := \sum_{i,j} |S_{ij}|.$$

⋆ *How about the existence of solution for the PCP problem?*
(cf. Candès-Li-Ma-Wright, J. ACM, 2011)

The penalty formulation and alternating direction method

Let $\mu > 0$ be the penalty parameter. Then we consider the relaxation using a penalty term to replace the constraint,

$$\min_{\boldsymbol{L},\boldsymbol{S}} \left(\|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S}\|_F^2 \right),$$

where $\|\cdot\|_F$ is the Frobenius norm. We set, for example, $\boldsymbol{S}^{(0)} = \mathbf{0}$. The ADM for the penalty formulation is given as follows: for $k \geq 0$, find

$$\begin{aligned} \boldsymbol{L}^{(k+1)} &= \arg\min_{\boldsymbol{L}} \left(\|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}^{(k)}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S}^{(k)}\|_F^2 \right), \\ \boldsymbol{S}^{(k+1)} &= \arg\min_{\boldsymbol{S}} \left(\|\boldsymbol{L}^{(k+1)}\|_* + \lambda \|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L}^{(k+1)} - \boldsymbol{S}\|_F^2 \right). \end{aligned}$$

By further analysis given below (pages 7-15), we can prove that

$$\begin{aligned} \boldsymbol{L}^{(k+1)} &= \mathrm{SVT}_{\frac{1}{\mu}} (\boldsymbol{M} - \boldsymbol{S}^{(k)}), \\ \boldsymbol{S}^{(k+1)} &= \mathrm{sign}(\boldsymbol{M} - \boldsymbol{L}^{(k+1)}) \odot \max \left\{ |\boldsymbol{M} - \boldsymbol{L}^{(k+1)}| - (\lambda/\mu),\ 0 \right\}, \end{aligned}$$

where $\odot$ is the Hadamard product (i.e., element-wise product).

SVD and SVT

- **Singular value decomposition (SVD):** *Let* $\boldsymbol{M} \in \mathbb{R}^{m \times n}$*. The SVD of* $\boldsymbol{M}$ *is the factorization in the form*

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top,$$

where $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ *and* $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ *are orthogonal matrices (*$\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{I}$ *and* $\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{I}$*) and* $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ *is diagonal with all non-negative entries called the singular values of* $\boldsymbol{M}$*.*

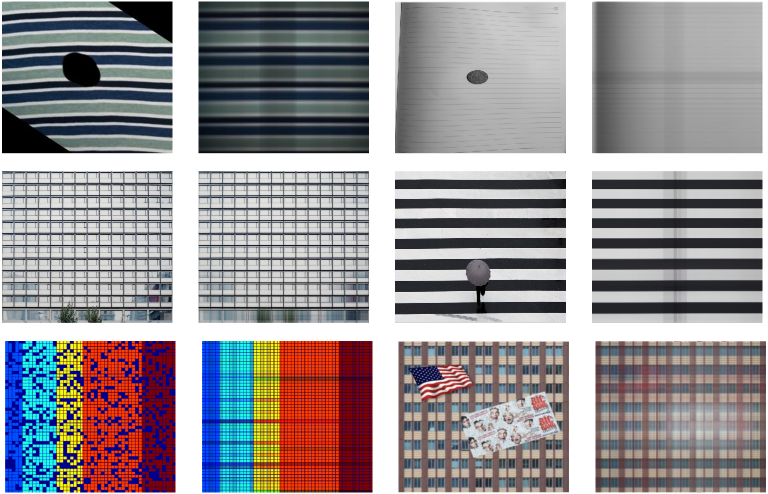
- **Singular value thresholding (SVT):** *Let* $\boldsymbol{M} \in \mathbb{R}^{m \times n}$*. Suppose that the SVD of* $\boldsymbol{M}$ *is given by* $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$*. Then the singular value thresholding of* $\boldsymbol{M}$ *with threshold* $\tau > 0$ *is defined by*

$$SVT_\tau(\boldsymbol{M}) = \boldsymbol{U}\boldsymbol{D}_\tau(\boldsymbol{\Sigma})\boldsymbol{V}^\top,$$

where

$$\boldsymbol{D}_\tau(\boldsymbol{\Sigma})_{ii} = \max\{\boldsymbol{\Sigma}_{ii} - \tau,\ 0\}, \quad \forall\, i = 1 : \min\{m, n\}.$$

Background recovering using the penalty method



Von Neumann trace inequality

First, we state without proof the square matrix case.

Theorem: *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are complex $n \times n$ matrices with singular values*

$$\sigma_1(\boldsymbol{A}) \geq \sigma_2(\boldsymbol{A}) \geq \cdots \geq \sigma_n(\boldsymbol{A}) \geq 0,$$
$$\sigma_1(\boldsymbol{B}) \geq \sigma_2(\boldsymbol{B}) \geq \cdots \geq \sigma_n(\boldsymbol{B}) \geq 0.$$

Then we have

$$|\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F| := |\text{trace}(\boldsymbol{A}^* \boldsymbol{B})| \leq \sum_{i=1}^{n} \sigma_i(\boldsymbol{A}) \sigma_i(\boldsymbol{B}).$$

Moreover, the equality holds if $\boldsymbol{A}$ and $\boldsymbol{B}$ share the same singular vectors.

Notes:

- If $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ then $\boldsymbol{A}^* = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^*$, having the same singular values $\sigma_i(\boldsymbol{A}^*) = \sigma_i(\boldsymbol{A}), \forall\, 1 \leq i \leq n.$ $\therefore$ $|\text{trace}(\boldsymbol{A}\boldsymbol{B})| \leq \sum_{i=1}^{n} \sigma_i(\boldsymbol{A}) \sigma_i(\boldsymbol{B})$.
- "Prove = if ...": If $\boldsymbol{A}$ and $\boldsymbol{B}$ share the same singular vectors, say $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}_A \boldsymbol{V}^*$ and $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}_B \boldsymbol{V}^*$, then we have
  $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{V}(\boldsymbol{\Sigma}_A \boldsymbol{\Sigma}_B)\boldsymbol{V}^* = \boldsymbol{V}(\boldsymbol{\Sigma}_B \boldsymbol{\Sigma}_A)\boldsymbol{V}^* = \boldsymbol{B}^*\boldsymbol{A} = (\boldsymbol{A}^*\boldsymbol{B})^*$, Hermitian!
  $\therefore$ $\text{trace}(\boldsymbol{A}^*\boldsymbol{B}) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{A}^*\boldsymbol{B}) = \sum_{i=1}^{n} \sigma_i(\boldsymbol{A}) \sigma_i(\boldsymbol{B}) \geq 0.$

Von Neumann trace inequality for rectangular matrices

Corollary: *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be complex $m \times n$ matrices with singular values*

$$\sigma_1(\boldsymbol{A}) \geq \sigma_2(\boldsymbol{A}) \geq \cdots \geq \sigma_k(\boldsymbol{A}) \geq 0,$$
$$\sigma_1(\boldsymbol{B}) \geq \sigma_2(\boldsymbol{B}) \geq \cdots \geq \sigma_k(\boldsymbol{B}) \geq 0,$$

where $k := \min\{m, n\}$. Then we have

$$|\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F| := |\mathrm{trace}(\boldsymbol{A}^* \boldsymbol{B})| \leq \sum_{i=1}^{k} \sigma_i(\boldsymbol{A}) \sigma_i(\boldsymbol{B}).$$

Moreover, the equality holds if $\boldsymbol{A}$ and $\boldsymbol{B}$ share the same singular vectors.

Proof: Assume that $m > n$. Then $k := \min\{m, n\} = n$. We define two $m \times m$ matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ by

$$\boldsymbol{X} = [\boldsymbol{A} \mid \mathbf{0}]_{m \times m} \quad \text{and} \quad \boldsymbol{Y} = [\boldsymbol{B} \mid \mathbf{0}]_{m \times m}.$$

Then we have

$$|\langle \boldsymbol{X}, \boldsymbol{Y} \rangle_F| = |\mathrm{trace}(\boldsymbol{X}^* \boldsymbol{Y})| = |\mathrm{trace}(\boldsymbol{A}^* \boldsymbol{B})| = |\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F|.$$

Proof of Von Neumann's trace inequality (cont'd)

Claim: $\sigma_i(\boldsymbol{X}) = \sigma_i(\boldsymbol{A})$ and similarly, $\sigma_i(\boldsymbol{Y}) = \sigma_i(\boldsymbol{B})$, $\forall\, i = 1, 2, \cdots, n$.

Suppose that the SVD of $\boldsymbol{A}$ is given by $\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}^*_{n\times n}$.
Define three $m \times m$ matrices,

$$\boldsymbol{U}_X = \boldsymbol{U}_{m\times m}, \quad \boldsymbol{\Sigma}_X = [\boldsymbol{\Sigma}_{m\times n} \mid \boldsymbol{0}]_{m\times m}, \quad \boldsymbol{V}_X^* = \begin{bmatrix} \boldsymbol{V}^*_{n\times n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}_{m\times m}.$$

Then we have

$$\begin{aligned} \boldsymbol{U}_X\boldsymbol{\Sigma}_X\boldsymbol{V}_X^* &= \boldsymbol{U}_{m\times m}[\boldsymbol{\Sigma}_{m\times n} \mid \boldsymbol{0}] \begin{bmatrix} \boldsymbol{V}^*_{n\times n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \\ &= [\boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n} \mid \boldsymbol{0}] \begin{bmatrix} \boldsymbol{V}^*_{n\times n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \\ &= [\boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\boldsymbol{V}^*_{n\times n} \mid \boldsymbol{0}] = [\boldsymbol{A}_{m\times n} \mid \boldsymbol{0}] = \boldsymbol{X}, \end{aligned}$$

which implies that $\sigma_i(\boldsymbol{X}) = \sigma_i(\boldsymbol{A})$, $\forall\, i = 1, 2, \cdots, n$. Therefore,

$$|\langle \boldsymbol{A}, \boldsymbol{B}\rangle_F| = |\langle \boldsymbol{X}, \boldsymbol{Y}\rangle_F| \le \sum_{i=1}^{n} \sigma_i(\boldsymbol{X})\sigma_i(\boldsymbol{Y}) = \sum_{i=1}^{n} \sigma_i(\boldsymbol{A})\sigma_i(\boldsymbol{B}). \quad \square$$

$SVT_\tau(\boldsymbol{Y})$ Theorem

Theorem: *Given an $m \times n$ real matrix $\boldsymbol{Y}$ and $\tau > 0$, we have*

$$SVT_\tau(\boldsymbol{Y}) = \arg\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \left(\tau \|\boldsymbol{X}\|_* + \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{Y}\|_F^2 \right).$$

Proof: Let $k := \min\{m, n\}$. Then for any $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, we have

$$\begin{aligned}
\frac{1}{2}\|\boldsymbol{X} - \boldsymbol{Y}\|_F^2 &= \frac{1}{2}\mathrm{tr}((\boldsymbol{X} - \boldsymbol{Y})^\top(\boldsymbol{X} - \boldsymbol{Y})) \\
&= \frac{1}{2}\mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{X}) - \mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{Y}) + \frac{1}{2}\mathrm{tr}(\boldsymbol{Y}^\top\boldsymbol{Y}) \\
&= \frac{1}{2}\sum_{i=1}^{n}\lambda_i(\boldsymbol{X}^\top\boldsymbol{X}) + \frac{1}{2}\sum_{i=1}^{n}\lambda_i(\boldsymbol{Y}^\top\boldsymbol{Y}) - \mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{Y}) \\
&\geq \frac{1}{2}\sum_{i=1}^{k}\sigma_i^2(\boldsymbol{X}) + \frac{1}{2}\sum_{i=1}^{k}\sigma_i^2(\boldsymbol{Y}) - \sum_{i=1}^{k}\sigma_i(\boldsymbol{X})\sigma_i(\boldsymbol{Y}) \\
&= \frac{1}{2}\sum_{i=1}^{k}\left(\sigma_i(\boldsymbol{X}) - \sigma_i(\boldsymbol{Y})\right)^2.
\end{aligned}$$

Proof of the $SVT_\tau(\boldsymbol{Y})$ Theorem (cont'd)

Therefore, we obtain for any $\boldsymbol{X} \in \mathbb{R}^{m\times n}$,

$$F(\boldsymbol{X}) := \tau\|\boldsymbol{X}\|_* + \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{Y}\|_F^2 \geq \tau\|\boldsymbol{X}\|_* + \frac{1}{2}\sum_{i=1}^{k}\big(\sigma_i(\boldsymbol{X})-\sigma_i(\boldsymbol{Y})\big)^2 =: G(\boldsymbol{X}).$$

It is already known that for a given $\tau > 0$ and a fixed $y \in \mathbb{R}$, the minimizer of the real-valued function,

$$f(x) = \tau|x| + \frac{1}{2}(y-x)^2, \quad x \in \mathbb{R},$$

is given by *the soft-thresholding operator* $\mathcal{S}_\tau$,

$$\arg\min_{x\in\mathbb{R}} f(x) = \mathcal{S}_\tau(y) := \operatorname{sign}(y)\max\{|y|-\tau,\ 0\}.$$

Also note that $\|\boldsymbol{X}\|_* = \sum_{i=1}^{k}\sigma_i(\boldsymbol{X})$. Therefore, we find the fact that

$$\begin{aligned}\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} G(\boldsymbol{X}) \quad \Leftrightarrow \quad \sigma_i(\widehat{\boldsymbol{X}}) &= \mathcal{S}_\tau(\sigma_i(\boldsymbol{Y})) \\ &= \operatorname{sign}(\sigma_i(\boldsymbol{Y}))\max\{|\sigma_i(\boldsymbol{Y})|-\tau,\ 0\} \\ &= \max\{\sigma_i(\boldsymbol{Y})-\tau,\ 0\},\ \forall\, i = 1,2,\cdots,k.\end{aligned}$$

$SVT_\tau(\boldsymbol{Y})$ Theorem (cont'd)

Based on the above observation, we are going to construct such a matrix $\widehat{\boldsymbol{X}}$ which has the same singular vectors with $\boldsymbol{Y}$. Suppose that the SVD of $\boldsymbol{Y}$ is given by $\boldsymbol{Y} = \boldsymbol{U\Sigma V}^\top$. Define the diagonal matrix $\widehat{\boldsymbol{\Sigma}}$ by

$$\widehat{\boldsymbol{\Sigma}} := \begin{bmatrix} \ddots & & \\ & \max\{\sigma_i(\boldsymbol{Y}) - \tau, 0\} & \\ & & \ddots \\ & & \\ & & \end{bmatrix}_{m\times n}$$

and then define $\widehat{\boldsymbol{X}} := \boldsymbol{U}\widehat{\boldsymbol{\Sigma}}\boldsymbol{V}^\top = SVT_\tau(\boldsymbol{Y})$. Therefore, *the equality in Von Neumann's trace inequality holds,* and we have

$$\tau\|\widehat{\boldsymbol{X}}\|_* + \frac{1}{2}\|\widehat{\boldsymbol{X}} - \boldsymbol{Y}\|_F^2 = \tau\|\widehat{\boldsymbol{X}}\|_* + \frac{1}{2}\sum_{i=1}^{k}\left(\sigma_i(\widehat{\boldsymbol{X}}) - \sigma_i(\boldsymbol{Y})\right)^2 = \min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} G(\boldsymbol{X}).$$

That is, we attain a minimum of $F(\boldsymbol{X})$ at $\widehat{\boldsymbol{X}} = SVT_\tau(\boldsymbol{Y})$.

$F(\boldsymbol{X})$ is a strictly convex function in $\boldsymbol{X} \in \mathbb{R}^{m\times n}$

Note that $F(\boldsymbol{X})$ is a strictly convex function in $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, since

- $\|\boldsymbol{X} - \boldsymbol{Y}\|_F^2$ is strictly convex in $\boldsymbol{X} \in \mathbb{R}^{m\times n}$.
- $\|\boldsymbol{X}\|_*$ is convex in $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, since it is a norm.
- "convex function + strictly convex function" is strictly convex.

Suppose that $\widehat{\boldsymbol{X}}_1$ and $\widehat{\boldsymbol{X}}_2$ are two different minimizers of the strictly convex function $F(\boldsymbol{X})$. Then

$$F(\frac{1}{2}(\widehat{\boldsymbol{X}}_1 + \widehat{\boldsymbol{X}}_2)) < \frac{1}{2}F(\widehat{\boldsymbol{X}}_1) + \frac{1}{2}F(\widehat{\boldsymbol{X}}_2) = F(\widehat{\boldsymbol{X}}_1), \text{ a contradiction!}$$

Therefore, the minimizer of $F(\boldsymbol{X})$ is unique! This completes the proof of the theorem. $\square$

Another direct proof of the uniqueness of minimizer $\widehat{X}$

Claim: *The minimizer of $F(\boldsymbol{X})$ is unique, that is, $\widehat{\boldsymbol{X}} = SVT_\tau(\boldsymbol{Y})$.*

Proof: Suppose that $\widehat{\boldsymbol{X}}_1$ and $\widehat{\boldsymbol{X}}_2$ are two different minimizers of $F(\boldsymbol{X})$. By the triangle inequality, we have

$$\tau\left\|\frac{\widehat{\boldsymbol{X}}_1+\widehat{\boldsymbol{X}}_2}{2}\right\|_* + \frac{1}{2}\left\|\frac{\widehat{\boldsymbol{X}}_1+\widehat{\boldsymbol{X}}_2}{2} - \boldsymbol{Y}\right\|_F^2$$

$$\leq \frac{\tau}{2}\|\widehat{\boldsymbol{X}}_1\|_* + \frac{\tau}{2}\|\widehat{\boldsymbol{X}}_2\|_* + \frac{1}{2}\left\|\frac{\widehat{\boldsymbol{X}}_1-\boldsymbol{Y}}{2} + \frac{\widehat{\boldsymbol{X}}_2-\boldsymbol{Y}}{2}\right\|_F^2. \quad (\star)$$

Note that

$$\left(\frac{a}{2}+\frac{b}{2}\right)^2 = \frac{a^2}{2}+\frac{b^2}{2} - \left(\frac{a-b}{2}\right)^2, \quad \forall\, a,b \in \mathbb{R}.$$

Therefore, we obtain

$$RHS(\star) = \frac{\tau}{2}\|\widehat{\boldsymbol{X}}_1\|_* + \frac{\tau}{2}\|\widehat{\boldsymbol{X}}_2\|_* + \frac{1}{4}\|\widehat{\boldsymbol{X}}_1-\boldsymbol{Y}\|_F^2 + \frac{1}{4}\|\widehat{\boldsymbol{X}}_2-\boldsymbol{Y}\|_F^2$$

$$-\frac{1}{2}\left\|\frac{\widehat{\boldsymbol{X}}_1-\widehat{\boldsymbol{X}}_2}{2}\right\|_F^2 = \tau\|\widehat{\boldsymbol{X}}_1\|_* + \frac{1}{2}\|\widehat{\boldsymbol{X}}_1-\boldsymbol{Y}\|_F^2 - \underbrace{\frac{1}{2}\left\|\frac{\widehat{\boldsymbol{X}}_1-\widehat{\boldsymbol{X}}_2}{2}\right\|_F^2}_{>0},$$

a contradiction!

Solution of the ADM for penalty formulation

By the $SVT_\tau(\boldsymbol{Y})$ Theorem, we have

$$\boldsymbol{L}^{(k+1)} := \arg\min_{\boldsymbol{L}} \left(\|\boldsymbol{L}\|_* + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S}^{(k)}\|_F^2 \right) = \mathrm{SVT}_{\frac{1}{\mu}} \left(\boldsymbol{M} - \boldsymbol{S}^{(k)} \right).$$

Using *the soft-thresholding operator* $\mathcal{S}_\tau$, we have

$$\begin{aligned} \boldsymbol{S}^{(k+1)} &:= \arg\min_{\boldsymbol{S}} \left(\lambda \|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L}^{(k+1)} - \boldsymbol{S}\|_F^2 \right) \\ &= \mathrm{sign}(\boldsymbol{M} - \boldsymbol{L}^{(k+1)}) \odot \max \left\{ |\boldsymbol{M} - \boldsymbol{L}^{(k+1)}| - (\lambda/\mu), \ 0 \right\}, \end{aligned}$$

where $\odot$ is the Hadamard element-wise product.

Another approach for solving the PCP problem

Recall the principal component pursuit problem:

$$\min_{\boldsymbol{L},\boldsymbol{S}} \left(\|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 \right) \quad \text{subject to} \quad \boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S}.$$

The augmented Lagrangian function is defined as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{L}, \boldsymbol{S}, \boldsymbol{Y}) \\ &:= \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 + \langle \underbrace{\boldsymbol{Y}}_{\textit{multiplier}}, \boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S} \rangle + \underbrace{\frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S}\|_F^2}_{\textit{penalty}} \\ &= \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S} + \mu^{-1} \boldsymbol{Y}\|_F^2 - \frac{1}{2\mu} \|\boldsymbol{Y}\|_F^2. \end{aligned}$$

We then apply the alternating direction method to minimize the augmented Lagrangian function $\mathcal{L}(\boldsymbol{L}, \boldsymbol{S}, \boldsymbol{Y})$. *The resulting method is called the augmented Lagrange multiplier (ALM) method. When* **$\boldsymbol{L}$** *and* **$\boldsymbol{S}$** *are further updated in an alternating way, it is also called the alternating direction method of multipliers (ADMM).*

The augmented Lagrange multiplier method

The ALM method is given by

$$
\begin{aligned}
\boldsymbol{L}^{(k+1)} &:= \arg\min_{\boldsymbol{L}} \Big(\|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}^{(k)}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{S}^{(k)} + \mu^{-1} \boldsymbol{Y}^{(k)}\|_F^2 \\
&\qquad - \frac{1}{2\mu} \|\boldsymbol{Y}^{(k)}\|_F^2 \Big), \\
\boldsymbol{S}^{(k+1)} &:= \arg\min_{\boldsymbol{S}} \Big(\|\boldsymbol{L}^{(k+1)}\|_* + \lambda \|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{M} - \boldsymbol{L}^{(k+1)} - \boldsymbol{S} + \mu^{-1} \boldsymbol{Y}^{(k)}\|_F^2 \\
&\qquad - \frac{1}{2\mu} \|\boldsymbol{Y}^{(k)}\|_F^2 \Big), \\
\boldsymbol{Y}^{(k+1)} &:= \boldsymbol{Y}^{(k)} + \mu \big(\boldsymbol{M} - \boldsymbol{L}^{(k+1)} - \boldsymbol{S}^{(k+1)} \big).
\end{aligned}
$$

The explicit form of the iterative solution $(\boldsymbol{L}^{(k+1)}, \boldsymbol{S}^{(k+1)}, \boldsymbol{Y}^{(k+1)})$ of ALM method is presented on the next page, which can be proved by using *the $SVT_\tau(\boldsymbol{Y})$ Theorem and the soft-thresholding operator S_τ.*

Iterative solutions of the ALM method

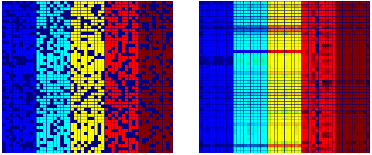
The iterative solution $(\boldsymbol{L}^{(k+1)}, \boldsymbol{S}^{(k+1)}, \boldsymbol{Y}^{(k+1)})$ of the ALM method is given by

$$
\begin{aligned}
\boldsymbol{L}^{(k+1)} &:= \arg\min_{\boldsymbol{L}} \Big(\|\boldsymbol{L}\|_* + \frac{\mu}{2} \|\boldsymbol{L} - (\boldsymbol{M} - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2 \Big) \\
&= \arg\min_{\boldsymbol{L}} \Big(\frac{1}{\mu} \|\boldsymbol{L}\|_* + \frac{1}{2} \|\boldsymbol{L} - (\boldsymbol{M} - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2 \Big) \\
&= SVT_{\frac{1}{\mu}} (\boldsymbol{M} - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)}), \\
\boldsymbol{S}^{(k+1)} &:= \arg\min_{\boldsymbol{S}} \Big(\lambda \|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{S} - (\boldsymbol{M} - \boldsymbol{L}^{(k+1)} + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2 \Big) \\
&= \arg\min_{\boldsymbol{S}} \Big(\frac{\lambda}{\mu} \|\boldsymbol{S}\|_1 + \frac{1}{2} \|\boldsymbol{S} - (\boldsymbol{M} - \boldsymbol{L}^{(k+1)} + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2 \Big) \\
&= \operatorname{sign}(\boldsymbol{M} - \boldsymbol{L}^{(k+1)} + \mu^{-1}\boldsymbol{Y}^{(k)}) \\
&\qquad \odot \max\left\{ |\boldsymbol{M} - \boldsymbol{L}^{(k+1)} + \mu^{-1}\boldsymbol{Y}^{(k)}| - (\lambda/\mu),\ 0 \right\}, \\
\boldsymbol{Y}^{(k+1)} &:= \boldsymbol{Y}^{(k)} + \mu (\boldsymbol{M} - \boldsymbol{L}^{(k+1)} - \boldsymbol{S}^{(k+1)}).
\end{aligned}
$$

Background recovering using the ALM method

$(\lambda, \mu) = (0.0007, 0.5)$

$(\lambda, \mu) = (0.006, 5)$



$(\lambda, \mu) = (0.007525, 0.04)$

$(\lambda, \mu) = (0.0025, 1.5)$

References

1. E. J. Candès, X. Li, Y. Ma, and J. Wright, Robust principal component analysis? *Journal of the ACM,* 58 (2011), Article 11.
2. X. Ren and Z. Lin, Linearized alternating direction method with adaptive penalty and warm starts for fast solving transform invariant low-rank textures, *International Journal of Computer Vision,* 104, (2013), pp.1-14.
3. Z. Lin, R. Liu, and Z. Su, Linearized alternating direction method with adaptive penalty for low-rank representation, *Proceedings of the 24th International Conference on Neural Information Processing Systems,* 2011, pp. 612-620.

Video foreground extraction

- Video foreground extraction is a fundamental task in computer vision, aiming to accurately extract the foreground from a sequence of frames containing *dynamic foreground objects and a static background.*
- Let $\boldsymbol{D} \in \mathbb{R}^{m\times n}$ be the data matrix formed by stacking vectorized grayscale video frames as columns of a video sequence. We assume that $\boldsymbol{D}$ is composed of a rank-1 background component and a sparse foreground component, modeled as

$$\boldsymbol{D} = \boldsymbol{u}\mathbf{1}^\top + \boldsymbol{S},$$

  where $\boldsymbol{u} \in \mathbb{R}^{m\times 1}$ and $\mathbf{1} \in \mathbb{R}^{n\times 1}$ is the all-ones vector.
- Based on this formulation, *we consider the following rank-1 principal component pursuit (PCP) problem:*

$$\min_{\boldsymbol{u}\in\mathbb{R}^{m\times 1}, \boldsymbol{S}\in\mathbb{R}^{m\times n}} \|\boldsymbol{S}\|_1 \quad \text{subject to } \boldsymbol{D} = \boldsymbol{u}\mathbf{1}^\top + \boldsymbol{S},$$

  where $\|\cdot\|_1$ denotes the ℓ_1 norm of $\boldsymbol{S}$ represented as a stacking long vector.

Rank-1 PCP problem

Define the augmented Lagrangian function by

$$\begin{aligned}\mathcal{L}(\boldsymbol{u}, \boldsymbol{S}, \boldsymbol{Y}) &:= \|\boldsymbol{S}\|_1 + \langle \boldsymbol{Y}, \boldsymbol{D} - \boldsymbol{u}\mathbf{1}^\top - \boldsymbol{S}\rangle + \frac{\mu}{2}\|\boldsymbol{D} - \boldsymbol{u}\mathbf{1}^\top - \boldsymbol{S}\|_F^2 \\ &= \|\boldsymbol{S}\|_1 + \frac{\mu}{2}\|\boldsymbol{D} - \boldsymbol{u}\mathbf{1}^\top - \boldsymbol{S} + \mu^{-1}\boldsymbol{Y}\|_F^2 - \frac{1}{2\mu}\|\boldsymbol{Y}\|_F^2,\end{aligned}$$

where $\mu > 0$ is a penalty parameter, $\boldsymbol{Y}$ is a Lagrange multiplier. The iterative scheme of the ALM method for the rank-1 PCP problem is as follows:

$$\begin{aligned}\boldsymbol{u}^{(k+1)} &:= \arg\min_{\boldsymbol{u}} \mathcal{L}_\mu(\boldsymbol{u}, \boldsymbol{S}^{(k)}, \boldsymbol{Y}^{(k)}) \\ &= \arg\min_{\boldsymbol{u}} \left(\|\boldsymbol{S}^{(k)}\|_1 + \frac{\mu}{2}\|\boldsymbol{D} - \boldsymbol{u}\mathbf{1}^\top - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)}\|_F^2 - \frac{1}{2\mu}\|\boldsymbol{Y}^{(k)}\|_F^2\right), \\ \boldsymbol{S}^{(k+1)} &:= \arg\min_{\boldsymbol{S}} \mathcal{L}_\mu(\boldsymbol{u}^{(k+1)}, \boldsymbol{S}, \boldsymbol{Y}^{(k)}) \\ &= \arg\min_{\boldsymbol{S}} \left(\|\boldsymbol{S}\|_1 + \frac{\mu}{2}\|\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top - \boldsymbol{S} + \mu^{-1}\boldsymbol{Y}^{(k)}\|_F^2 - \frac{1}{2\mu}\|\boldsymbol{Y}^{(k)}\|_F^2\right), \\ \boldsymbol{Y}^{(k+1)} &:= \boldsymbol{Y}^{(k)} + \mu(\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top - \boldsymbol{S}^{(k+1)}).\end{aligned}$$

The closed-form solutions

The iterative update $(\boldsymbol{u}^{(k+1)}, \boldsymbol{S}^{(k+1)}, \boldsymbol{Y}^{(k+1)})$ has closed-form solutions. With $\boldsymbol{S}^{(k)}$ and $\boldsymbol{Y}^{(k)}$ fixed, the update for $\boldsymbol{u}^{(k+1)}$ is obtained by solving

$$\begin{aligned}\boldsymbol{u}^{(k+1)} &:= \arg\min_{\boldsymbol{u}} \left(\frac{\mu}{2}\|\boldsymbol{u}\mathbf{1}^\top - (\boldsymbol{D} - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2\right) \\ &= \arg\min_{\boldsymbol{u}} \left(\frac{\mu}{2}\|\boldsymbol{u}\mathbf{1}^\top - \boldsymbol{W}\|_F^2\right),\end{aligned}$$

where $\boldsymbol{W} := \boldsymbol{D} - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)}$. This subproblem can be solved analytically as follows:

$$\begin{aligned}\boldsymbol{u}^{(k+1)} &= \arg\min_{\boldsymbol{u}} \left(\langle \boldsymbol{W}, \boldsymbol{W}\rangle - 2\langle \boldsymbol{W}, \boldsymbol{u}\mathbf{1}^\top\rangle + \langle \boldsymbol{u}\mathbf{1}^\top, \boldsymbol{u}\mathbf{1}^\top\rangle\right) \\ &= \arg\min_{\boldsymbol{u}} \left(\|\boldsymbol{W}\|_F^2 - 2\boldsymbol{u}^\top \boldsymbol{W}\mathbf{1} + n\|\boldsymbol{u}\|_2^2\right) \\ &= \arg\min_{\boldsymbol{u}} \left(-2\boldsymbol{u}^\top \boldsymbol{W}\mathbf{1} + n\|\boldsymbol{u}\|_2^2\right) \\ &= \frac{1}{n}(\boldsymbol{D} - \boldsymbol{S}^{(k)} + \mu^{-1}\boldsymbol{Y}^{(k)}) \cdot \mathbf{1}.\end{aligned}$$

The closed-form solutions (cont'd)

With $\boldsymbol{u}^{(k+1)}$ fixed, the variable $\boldsymbol{S}^{(k+1)}$ is updated by solving

$$\begin{aligned} \boldsymbol{S}^{(k+1)} &:= \arg\min_{\boldsymbol{S}} \left(\|\boldsymbol{S}\|_1 + \frac{\mu}{2} \|\boldsymbol{S} - (\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2 \right) \\ &= \arg\min_{\boldsymbol{S}} \left(\frac{1}{\mu} \|\boldsymbol{S}\|_1 + \frac{1}{2} \|\boldsymbol{S} - (\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top + \mu^{-1}\boldsymbol{Y}^{(k)})\|_F^2 \right). \end{aligned}$$

This subproblem yields a component-wise soft-thresholding solution:

$$\begin{aligned} \boldsymbol{S}^{(k+1)} = \operatorname{sign}(\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top + \mu^{-1}\boldsymbol{Y}^{(k)}) \\ \odot \max\{|\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top + \mu^{-1}\boldsymbol{Y}^{(k)}| - (1/\mu),\ 0\}, \end{aligned}$$

and $\odot$ denotes the Hadamard (element-wise) product.

Finally, the Lagrange multiplier is updated as

$$\boldsymbol{Y}^{(k+1)} := \boldsymbol{Y}^{(k)} + \mu(\boldsymbol{D} - \boldsymbol{u}^{(k+1)}\mathbf{1}^\top - \boldsymbol{S}^{(k+1)}).$$

Video foreground extraction

- **[Wei-2025]** S.-T. Wei, Video foreground extraction via efficient rank-1 principal component pursuit, *Master Thesis,* Department of Mathematics, National Central University, June 2025.
- **[MA3111]** S.-Y. Yang, Principal component pursuit problem, *Lecture Slides,* 2025.