Principal Component Pursuit and Transform Invariant Low-Rank Textures

Suh-Yuh Yang (楊肅煜)

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

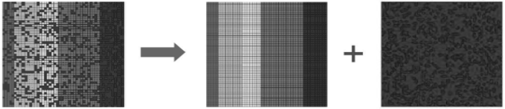
First version: September 30, 2022/Last updated: March 12, 2024

Sparse plus low rank matrix decomposition

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

$$\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S}.$$

We are interested in finding the low-rank image $\boldsymbol{L}$, 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:

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

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

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* $\mathbb{R}^{mn}$*)*,

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

⋆ *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, $\boldsymbol{S}^{(0)} = \boldsymbol{0}$. 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 (pp. 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*

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

  where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ *and* $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ *are orthogonal matrices* ($\boldsymbol{UU}^\top = \boldsymbol{I}$ *and* $\boldsymbol{VV}^\top = \boldsymbol{I}$) *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)**

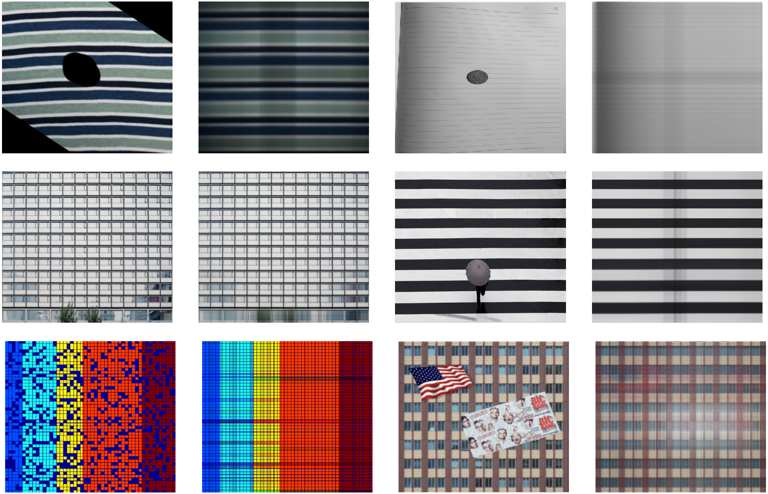
  Let $\boldsymbol{M} \in \mathbb{R}^{m\times n}$. *Suppose that the SVD of* $\boldsymbol{M}$ *is given by* $\boldsymbol{M} = \boldsymbol{U\Sigma V}^\top$. *Then the singular value thresholding (SVT) of* $\boldsymbol{M}$ *with threshold* $\tau > 0$ *is defined by*

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

  where

$$\boldsymbol{D}_\tau(\boldsymbol{\Sigma})_{ii} = \max\{\boldsymbol{\Sigma}_{ii} - \tau,\ 0\}.$$

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

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

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

Notes:

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

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

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

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

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

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

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

$SVT_\tau(\boldsymbol{Y})$ 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 $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$. 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

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

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

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

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

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

$$RHS(\star) = \frac{\tau}{2} \| \widehat{\boldsymbol{X}}_1 \|_* + \frac{\tau}{2} \| \widehat{\boldsymbol{X}}_2 \|_* + \frac{1}{4} \| \widehat{\boldsymbol{X}}_1 - \boldsymbol{Y} \|_F^2 + \frac{1}{4} \| \widehat{\boldsymbol{X}}_2 - \boldsymbol{Y} \|_F^2$$
$$- \frac{1}{2} \| \frac{\widehat{\boldsymbol{X}}_1 - \widehat{\boldsymbol{X}}_2}{2} \|_F^2 = \tau \| \widehat{\boldsymbol{X}}_1 \|_* + \frac{1}{2} \| \widehat{\boldsymbol{X}}_1 - \boldsymbol{Y} \|_F^2 - \underbrace{\frac{1}{2} \| \frac{\widehat{\boldsymbol{X}}_1 - \widehat{\boldsymbol{X}}_2}{2} \|_F^2}_{>0},$$

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* $\mathcal{S}_\tau$ again, 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:

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

The augmented Lagrangian function is defined as

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

We then apply the alternating direction method to minimize the augmented Lagrangian function $\mathcal{L}(L, S, Y)$. The resulting method is called *the augmented Lagrange multiplier (ALM) method.*

The augmented Lagrange multiplier method

The ALM method is given by

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

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

Iterative solutions of the ALM method

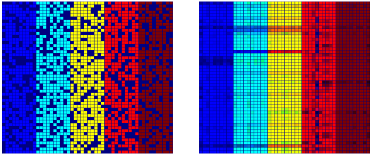
The iterative solution $(\boldsymbol{L}^{(k+1)}, \boldsymbol{S}^{(k+1)}, \boldsymbol{Y}^{(k+1)})$ of the ALM method is given by

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

Background recovering using the ALM method

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

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



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

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

Low-rank textures

Consider a 2D texture as a matrix $\boldsymbol{L} \in \mathbb{R}^{p \times q}$. It is called a low-rank texture if $r := \operatorname{rank}(\boldsymbol{L}) \ll \min\{p, q\}$.

- A real texture image is hardly an ideal low-rank texture, mainly due to two factors
  (1) *It undergoes a deformation, e.g., a perspective transform from 3D scene to 2D image;*
  (2) *It may be subject to many types of corruption, such as noise and occlusion.*

- Suppose that a *larger* low-rank texture $\boldsymbol{L}$ lies on a planar surface in the scene. The *smaller* $m \times n$ image $\boldsymbol{M}$ that we observe from a certain viewpoint is a portion of the transformed version of $\boldsymbol{L}$. Then there exists an *invertible function* $\boldsymbol{\tau}^{-1} : \mathbb{N}^2 \to \mathbb{N}^2$ such that

$$\boldsymbol{M}(i,j) = (\boldsymbol{L} \circ \boldsymbol{\tau}^{-1})(i,j) = \boldsymbol{L}(\boldsymbol{\tau}^{-1}(i,j)), \quad \forall\, (i,j) \in K,$$

where $K := \{(i,j) \in \mathbb{N}^2 : 1 \le i \le m, 1 \le j \le n\}$.

Transform invariant low-rank textures (TILT)

In addition to domain transformations, the observed image of the texture might be corrupted by noise and occlusions, denoted as $\boldsymbol{S}$.

- Then we have

$$\boldsymbol{M}(i,j) = ((\boldsymbol{L}+\boldsymbol{S})\circ\boldsymbol{\tau}^{-1})(i,j), \quad \forall\,(i,j)\in K.$$

That is,

$$(\boldsymbol{M}\circ\boldsymbol{\tau})(i,j) = \boldsymbol{L}(i,j)+\boldsymbol{S}(i,j), \quad \forall\,(i,j)\in K,$$

- A typical perspective transform from 3D scene to 2D image is the affine transformation, i.e.,

$$\boldsymbol{\tau}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}, \quad \boldsymbol{x}\in\mathbb{R}^2,$$

where $\boldsymbol{A}\in\mathbb{R}^{2\times 2}$ is an *invertible matrix* and $\boldsymbol{b}\in\mathbb{R}^2$ is a constant vector.

The mathematical model for TILT

So, if we could rectify a deformed texture $\boldsymbol{M}$ with a proper inverse transform $\boldsymbol{\tau}$ and then remove the corruptions $\boldsymbol{S}$, the resulting texture $\boldsymbol{L}$ will be low rank. The mathematical model for TILT is given by

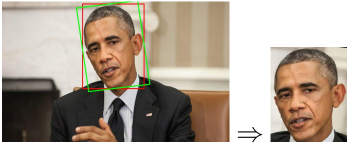
$$\min_{\boldsymbol{L},\boldsymbol{S},\boldsymbol{\tau}} \left(\operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{S}\|_0\right) \quad \text{subject to} \quad \boldsymbol{M} \circ \boldsymbol{\tau} = \boldsymbol{L} + \boldsymbol{S}.$$

In practice, the rank and the ℓ^0-norm could be replaced by the nuclear norm and ℓ^1-norm, respectively:

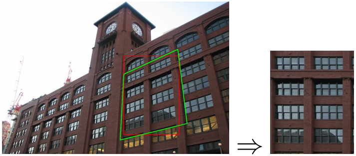
$$\min_{\boldsymbol{L},\boldsymbol{S},\boldsymbol{\tau}} \left(\|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{S}\|_1\right) \quad \text{subject to} \quad \boldsymbol{M} \circ \boldsymbol{\tau} = \boldsymbol{L} + \boldsymbol{S},$$

where the constraint is non-convex. *Therefore, we have to consider the linearization of* $\boldsymbol{M} \circ \boldsymbol{\tau}$.

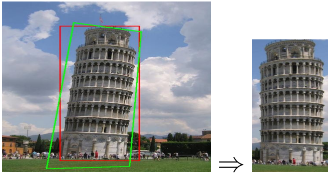
Numerical examples of TILT



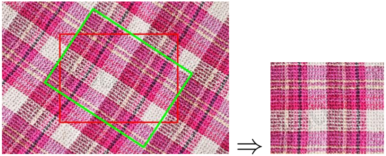
A: $(\lambda, \mu) = (1/257, 3.1672\text{e-}5)$



H: $(\lambda, \mu) = (1/387, 5.0743\text{e-}5)$



H: $(\lambda, \mu) = (1/505, 3.7585\text{e-}5)$



A: $(\lambda, \mu) = (1/186, 2.8748\text{e-}5)$

A class of convex minimization problems

We consider the following convex minimization problems where the objective function is separable:

$$\min_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \quad \text{subject to } \mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}) = \boldsymbol{c},$$

where f and g are convex real-valued functions, $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{c}$ could be either vectors or matrices, and $\mathcal{A}$ and $\mathcal{B}$ are linear mappings.

Define the augmented Lagrangian function

$$\begin{aligned} \mathcal{L}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) &:= f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}, \mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}) - \boldsymbol{c} \rangle \\ &\quad + \frac{\beta}{2} \|\mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}) - \boldsymbol{c}\|_F^2 \\ &= f(\boldsymbol{x}) + g(\boldsymbol{y}) + \frac{\beta}{2} \|\mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}) - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}\|_F^2 - \frac{1}{2\beta}\|\boldsymbol{\lambda}\|_F^2, \end{aligned}$$

where $\boldsymbol{\lambda}$ is the Lagrange multiplier, $\langle \cdot, \cdot \rangle$ is the inner product, and $\beta > 0$ is the penalty parameter.

The augmented Lagrange multiplier method

- We apply the alternating direction method to minimize the function $\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$. The resulting ALM method decomposes the minimization of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$ w.r.t. $(\boldsymbol{x}, \boldsymbol{y})$ into two subproblems:

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \arg\min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}^{(k)}, \boldsymbol{\lambda}^{(k)}), \\
\boldsymbol{y}^{(k+1)} &= \arg\min_{\boldsymbol{y}} \mathcal{L}(\boldsymbol{x}^{(k+1)}, \boldsymbol{y}, \boldsymbol{\lambda}^{(k)}), \\
\boldsymbol{\lambda}^{(k+1)} &= \boldsymbol{\lambda}^{(k)} + \beta\big(\mathcal{A}(\boldsymbol{x}^{(k+1)}) + \mathcal{B}(\boldsymbol{y}^{(k+1)}) - \boldsymbol{c}\big).
\end{aligned}
$$

- In compressive sensing and sparse representation, as f and g are usually matrix or vector norms, the first two subproblems usually have closed form solutions when $\mathcal{A}$ and $\mathcal{B}$ are identities.

Linearized alternating direction method

However, in many problems $\mathcal{A}$ and $\mathcal{B}$ are not identities, we consider a linearization technique. First, we focus on the first subproblem which can be rewritten as

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{y}^{(k)}) + \langle \boldsymbol{\lambda}^{(k)}, \mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}^{(k)}) - \boldsymbol{c}\rangle \\
&\qquad + \frac{\beta}{2}\|\mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}^{(k)}) - \boldsymbol{c}\|_F^2 \\
&= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{y}^{(k)}) \\
&\qquad + \frac{\beta}{2}\|\mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}^{(k)}) - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}^{(k)}\|_F^2 - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{(k)}\|_F^2 \\
&= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\beta}{2}\|\mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{y}^{(k)}) - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}^{(k)}\|_F^2.
\end{aligned}
$$

Linearized alternating direction method (cont'd)

Now, we have

$$x^{(k+1)} = \arg\min_{x} f(x) + \frac{\beta}{2}\|\mathcal{A}(x) + \mathcal{B}(y^{(k)}) - c + \frac{1}{\beta}\lambda^{(k)}\|_F^2.$$

Define $H(x)$ as the quadratic function,

$$H(x) = \|\mathcal{A}(x) + \mathcal{B}(y^{(k)}) - c + \frac{1}{\beta}\lambda^{(k)}\|_F^2.$$

By the Taylor expansion at $x^{(k)}$, we have

$$H(x) \approx H(x^{(k)}) + \nabla H(x^{(k)}) \cdot (x - x^{(k)}).$$

Then the minimization problem approximately becomes

$$x^{(k+1)} = \arg\min_{x} f(x) + \frac{\beta}{2}H(x^{(k)}) + \frac{\beta}{2}\nabla H(x^{(k)}) \cdot (x - x^{(k)}).$$

Linearized alternating direction method (cont'd)

In what follows, we assume that $\mathcal{A}$ and $\mathcal{B}$ are real matrices. Assume that $\mathcal{A} = [a_{ij}]_{m\times n} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} = (b_1, b_2, \cdots, b_m)^\top \in \mathbb{R}^m$. Then

$$\begin{aligned} \nabla_{\boldsymbol{x}} \langle \mathcal{A}\boldsymbol{x}, \boldsymbol{b}\rangle &= \nabla \Big(b_1(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + \\ &\qquad b_2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n) + \cdots + \\ &\qquad b_m(a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n)\Big) \\ &= (\mathcal{A}_{\cdot\,1} \cdot \boldsymbol{b}, \mathcal{A}_{\cdot\,2} \cdot \boldsymbol{b}, \cdots, \mathcal{A}_{\cdot\,n} \cdot \boldsymbol{b})^\top = \mathcal{A}^\top \boldsymbol{b}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \nabla H(\boldsymbol{x}) &= \nabla_{\boldsymbol{x}} \langle \mathcal{A}\boldsymbol{x} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}^{(k)},\ \mathcal{A}\boldsymbol{x} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}^{(k)} \rangle \\ &= 2\mathcal{A}^\top \Big(\mathcal{A}\boldsymbol{x} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}^{(k)}\Big). \end{aligned}$$

Linearized alternating direction method (cont'd)

Using the linearization of $H(\boldsymbol{x})$ at $\boldsymbol{x}^{(k)}$ and adding a proximal term, which ensures the Taylor approximation reasonable since $\boldsymbol{x}$ close to $\boldsymbol{x}^{(k)}$, we have the following approximation:

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \beta\langle \mathcal{A}^\top(\mathcal{A}\boldsymbol{x}^{(k)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c} + \frac{1}{\beta}\boldsymbol{\lambda}^{(k)}),\ \boldsymbol{x} - \boldsymbol{x}^{(k)}\rangle + \frac{\beta\eta_{\mathcal{A}}}{2}\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|_F^2 \\
&= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \langle \mathcal{A}^\top\boldsymbol{\lambda}^{(k)} + \beta\mathcal{A}^\top(\mathcal{A}\boldsymbol{x}^{(k)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}),\ \boldsymbol{x} - \boldsymbol{x}^{(k)}\rangle + \frac{\beta\eta_{\mathcal{A}}}{2}\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|_F^2 \\
&= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\beta\eta_{\mathcal{A}}}{2}\|(\boldsymbol{x} - \boldsymbol{x}^{(k)}) + \mathcal{A}^\top(\boldsymbol{\lambda}^{(k)} + \beta(\mathcal{A}\boldsymbol{x}^{(k)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))/(\beta\eta_{\mathcal{A}})\|_F^2 \\
&\qquad - \frac{1}{2\beta\eta_{\mathcal{A}}}\|\mathcal{A}^\top(\boldsymbol{\lambda}^{(k)} + \beta(\mathcal{A}\boldsymbol{x}^{(k)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))\|_F^2,
\end{aligned}
$$

where $\eta_{\mathcal{A}} > \|\mathcal{A}\|_F^2 > 0$ is a parameter in the proximal term. Similarly, the second subproblem can be approximated by

$$
\boldsymbol{y}^{(k+1)} = \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \frac{\beta\eta_{\mathcal{B}}}{2}\|(\boldsymbol{y} - \boldsymbol{y}^{(k)}) + \mathcal{B}^\top(\boldsymbol{\lambda}^{(k)} + \beta(\mathcal{A}\boldsymbol{x}^{(k+1)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))/(\beta\eta_{\mathcal{B}})\|_F^2,
$$

where $\eta_{\mathcal{B}} > \|\mathcal{B}\|_F^2 > 0$ is a parameter in the proximal term.

Linearized alternating direction method (cont'd)

To sum up, the linearized alternating direction method is given by

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\beta\eta_{\mathcal{A}}}{2}\|(\boldsymbol{x}-\boldsymbol{x}^{(k)}) + \mathcal{A}^\top(\boldsymbol{\lambda}^{(k)} + \beta(\mathcal{A}\boldsymbol{x}^{(k)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))/(\beta\eta_{\mathcal{A}})\|_F^2, \\
\boldsymbol{y}^{(k+1)} &= \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \frac{\beta\eta_{\mathcal{B}}}{2}\|(\boldsymbol{y}-\boldsymbol{y}^{(k)}) + \mathcal{B}^\top(\boldsymbol{\lambda}^{(k)} + \beta(\mathcal{A}\boldsymbol{x}^{(k+1)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))/(\beta\eta_{\mathcal{B}})\|_F^2, \\
\boldsymbol{\lambda}^{(k+1)} &= \boldsymbol{\lambda}^{(k)} + \beta(\mathcal{A}\boldsymbol{x}^{(k+1)} + \mathcal{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{c}),
\end{aligned}$$

with one of the stopping criteria or both:

- First stopping criterion:

$$\|\mathcal{A}\boldsymbol{x}^{(k+1)} + \mathcal{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{c}\|_F < \epsilon_1\|\boldsymbol{c}\|_F.$$

- Second stopping criterion:

$$\beta\max\left(\sqrt{\eta_{\mathcal{A}}}\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_F,\ \sqrt{\eta_{\mathcal{B}}}\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^{(k)}\|_F\right) < \epsilon_2\|\boldsymbol{c}\|_F.$$

Linearized ADM with adaptive penalty (LADMAP)

To further accelerate the convergence of the algorithm, we also consider an adaptive rule for updating β. Consider the following adaptive updating strategy for the penalty parameter:

$$\beta_{k+1} = \min(\beta_{\max}, \rho\beta_k),$$

where $\beta_{\max}$ is an upper bound of $\{\beta_k\}$ and ρ is defined as

$$\rho = \begin{cases} \rho_0, & \text{if } \frac{\beta_k}{\|\boldsymbol{c}\|_F} \max\left(\sqrt{\eta_{\mathcal{A}}}\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_F, \sqrt{\eta_{\mathcal{B}}}\|\boldsymbol{y}^{(k+1)} - \boldsymbol{y}^{(k)}\|_F\right) < \epsilon_2, \\ 1, & \text{otherwise,} \end{cases}$$

and $\rho_0 > 1$ is a constant. The LADMAP is defined as

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\beta_k\eta_{\mathcal{A}}}{2}\|(\boldsymbol{x} - \boldsymbol{x}^{(k)}) + \mathcal{A}^\top(\boldsymbol{\lambda}^{(k)} + \beta_k(\mathcal{A}\boldsymbol{x}^{(k)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))/(\beta_k\eta_{\mathcal{A}})\|_F^2, \\ \boldsymbol{y}^{(k+1)} &= \arg\min_{\boldsymbol{y}} g(\boldsymbol{y}) + \frac{\beta_k\eta_{\mathcal{B}}}{2}\|(\boldsymbol{y} - \boldsymbol{y}^{(k)}) + \mathcal{B}^\top(\boldsymbol{\lambda}^{(k)} + \beta_k(\mathcal{A}\boldsymbol{x}^{(k+1)} + \mathcal{B}\boldsymbol{y}^{(k)} - \boldsymbol{c}))/(\beta_k\eta_{\mathcal{B}})\|_F^2, \\ \boldsymbol{\lambda}^{(k+1)} &= \boldsymbol{\lambda}^{(k)} + \beta_k(\mathcal{A}\boldsymbol{x}^{(k+1)} + \mathcal{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{c}), \\ \beta_{k+1} &= \min(\beta_{\max}, \rho\beta_k). \end{aligned}$$

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.
4. G. Tang and A. Nehorai, Robust PCA based on low-rank and block-sparse matrix decomposition, *45th Annual Conference on Information Sciences and Systems,* 2011, pp. 1-5.
5. Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, TILT: transform invariant low-rank textures, *International Journal of Computer Vision,* 99 (2012), pp. 1-24.