Positive Collaborative Representation for Subspace Clustering

June 10, 2017

Abstract

1 Introduction

2 Motivation

Motivation 1: Positive collaborative representation could achieve sparse representation since similar points are sparse while dissimilar points are dense.

Motivation 2: Positive supports are positive to self-reppresentation while negative supports are negative to self-representation.

Motivation 3: Better performance. Faster?

3 LSR Model

The least squares regression (LSR) model [] is proposed by Lu et al. can be formulated as follows:

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2, \tag{1}$$

where $0 \leq p \leq 1$. For analysis simplicity, we just set $p = 1$.

Then the model above can be

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2. \tag{2}$$

By introducing auxiliary variables into the optimization program, we can set $\boldsymbol{E} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}$. The program (42) can be transformed into

$$\min_{\boldsymbol{A},\boldsymbol{E}} \|\boldsymbol{E}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2 \text{ s.t. } \boldsymbol{E} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}. \tag{3}$$

By introducing a Lagrangian multiplier $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (15) can be written as

$$\mathcal{L}(\boldsymbol{A}, \boldsymbol{E}, \boldsymbol{\Delta}, \rho) = \|\boldsymbol{E}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2$$
$$+ \frac{\rho}{2}\|\boldsymbol{E} - \boldsymbol{X} + \boldsymbol{X}\boldsymbol{A} + \rho^{-1}\boldsymbol{\Delta}\|_F^2 \tag{4}$$

Denote by $(\boldsymbol{A}_k, \boldsymbol{E}_k)$ the optimization variables at iteration k, and by $\boldsymbol{\Delta}$ the Lagrangian multipliers at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing $(\boldsymbol{E}_k, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\boldsymbol{A}_{k+1} = \arg\min_{\boldsymbol{A}} \frac{\rho_k}{2}\|\boldsymbol{P}_k - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_F^2 \tag{5}$$

where $\boldsymbol{P}_k = \boldsymbol{X} - \boldsymbol{E}_k - \rho_k^{-1}\boldsymbol{\Delta}_k$. This is a least squares regression problem which has a closed-form solution as

$$\boldsymbol{A}_{k+1} = (\boldsymbol{X}^\top\boldsymbol{X} + 2\lambda\rho_k^{-1}\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{P}_k. \tag{6}$$

(3) Obtain $\boldsymbol{E}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{E}$, while fixing $(\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\min_{\boldsymbol{E}} \frac{\rho_k}{2}\|\boldsymbol{Q}_k - \boldsymbol{E}\|_F^2 + \|\boldsymbol{E}\|_p^p, \tag{7}$$

where $\boldsymbol{Q}_k = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}_{k+1} - \rho_k^{-1}\boldsymbol{\Delta}_k$. Since this problem is seperable, it can be decoupled into $m \times n$ scalable sub-problems which share the following form:

$$\min_y = \frac{1}{2}(c - y)^2 + \rho_k^{-1}|y|^p. \tag{8}$$

This problem can be solved by the generalized soft-thresholding algorithm [**?**] as

$$\hat{y} = \text{sign}(c) * \max(|c| - \tau_p(\rho_k^{-1}), 0) * \mathcal{S}_p(|c|; \rho_k^{-1}), \tag{9}$$

where

$$\tau_p(\rho_k^{-1}) = [2\rho_k^{-1}(1-p)]^{\frac{1}{2-p}} + \rho_k^{-1}p[2\rho_k^{-1}(1-p)]^{\frac{p-1}{2-p}} \tag{10}$$

and $\mathcal{S}_p(|c|; \rho_k^{-1})$ can be obtained by solving the following equation:

$$\mathcal{S}_p(|c|; \rho_k^{-1}) - c + \rho_k^{-1}p(\mathcal{S}_p(|c|; \rho_k^{-1}))^{p-1} = 0. \tag{11}$$

(4) Obtain the Lagrangian multipliers ($\boldsymbol{\Delta}_{k+1}$) while fixing ($\boldsymbol{A}_{k+1}, \boldsymbol{E}_{k+1}$):

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \tau\rho_k(\boldsymbol{E}_{k+1} - \boldsymbol{X} + \boldsymbol{X}\boldsymbol{A}_{k+1}), \tag{12}$$

where $\tau \in (0, \frac{\sqrt{5}+1}{2})$ is the dual step size and is usually set as $\tau = 1$.

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu\rho_k$, where $\mu > 1$.

4 RES with Diagonal Constraint

We can also restrict that $\text{diag}(\boldsymbol{A}) = \boldsymbol{0}$ to avoid the samples to be self-represented. However, I want to mention that the proposed model solved by ADMM algorithm with three variables and does not have convergence results.

Then the model above can be

$$\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2 \quad \text{s.t.} \quad \text{diag}(\boldsymbol{A}) = \boldsymbol{0}. \tag{13}$$

By introducing auxiliary variables into the optimization program, we can set $\boldsymbol{E} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}$ and $\boldsymbol{A} = \boldsymbol{0}$. The program (42) can be transformed into

$$\min_{\boldsymbol{A},\boldsymbol{E}} \|\boldsymbol{E}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2$$
$$\text{s.t. } \boldsymbol{E} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}, \text{diag}(\boldsymbol{A}) = \boldsymbol{0}. \tag{14}$$

Besides, we can further introduce an auxiliary matrix variable $\boldsymbol{C}$ and consider the following program

$$\min_{\boldsymbol{A},\boldsymbol{C},\boldsymbol{E}} \|\boldsymbol{E}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2$$
$$\text{s.t. } \boldsymbol{E} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}, \boldsymbol{C} = \boldsymbol{A} - \text{diag}(\boldsymbol{A}) \tag{15}$$

whose solution for ($\boldsymbol{A}, \boldsymbol{E}$) coincides with the solution of Eq. (43). By introducing two Lagrangian multipliers

$\boldsymbol{\Delta}, \boldsymbol{\delta}$, the Lagrangian function of the Eq. (15) can be written as

$$\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{E}, \boldsymbol{\Delta}, \boldsymbol{\delta}, \rho) = \|\boldsymbol{E}\|_p^p + \lambda\|\boldsymbol{A}\|_F^2$$
$$+ \frac{\rho}{2}\|\boldsymbol{E} - \boldsymbol{X} + \boldsymbol{X}\boldsymbol{C} + \rho^{-1}\boldsymbol{\Delta}\|_F^2 \tag{16}$$
$$+ \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A} + \text{diag}(\boldsymbol{A}) + \rho^{-1}\boldsymbol{\delta}\|_F^2$$

Denote by ($\boldsymbol{C}_k, \boldsymbol{A}_k, \boldsymbol{E}_k$) the optimization variables at iteration k, by ($\boldsymbol{\Delta}_k, \boldsymbol{\delta}_k$) the Lagrangian multipliers at iteration k, and by ρ_k the penalty parameter at iteration k. Taking derivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing ($\boldsymbol{C}_k, \boldsymbol{E}_k, \boldsymbol{\Delta}_k, \boldsymbol{\delta}_k$). This is equivalent to solve the following problem:

$$\boldsymbol{A}_{k+1} = \boldsymbol{J} - \text{diag}(\boldsymbol{J}),$$
$$\boldsymbol{J} = (\rho_k + 2\lambda)^{-1}(\rho_k\boldsymbol{C}_k + \boldsymbol{\delta}_k) \tag{17}$$

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing ($\boldsymbol{A}_{k+1}, \boldsymbol{E}_k, \boldsymbol{\Delta}_k, \boldsymbol{\delta}_k$). This is equivalent to solve the following problem:

$$\boldsymbol{C}_{k+1} = \arg\min_{\boldsymbol{C}} \frac{\rho_k}{2}\|\boldsymbol{E}_k - \boldsymbol{X} + \boldsymbol{X}\boldsymbol{C} + \rho_k^{-1}\boldsymbol{\Delta}_k\|_F^2$$
$$+ \frac{\rho_k}{2}\|\boldsymbol{C} - \boldsymbol{A}_{k+1} + \rho_k^{-1}\boldsymbol{\delta}_k\|_F^2 \tag{18}$$

This is a least squares regression problem which has a closed-form solution as

$$\boldsymbol{C}_{k+1} = \boldsymbol{H} - \text{diag}(\boldsymbol{H}),$$
$$(\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{I})\boldsymbol{H} = \boldsymbol{X}^\top\boldsymbol{P}_k + \boldsymbol{Q}_k, \tag{19}$$

where $\boldsymbol{P}_k = \boldsymbol{X} - \boldsymbol{E}_k - \rho_k^{-1}\boldsymbol{\Delta}_k$ and $\boldsymbol{Q}_k = \boldsymbol{A}_{k+1} - \rho_k^{-1}\boldsymbol{\delta}_k$.

(3) Obtain $\boldsymbol{E}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{E}$, while fixing ($\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k, \boldsymbol{\delta}_k$). This is equivalent to solve the following problem:

$$\min_{\boldsymbol{E}} \frac{1}{2}\|(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}_{k+1} - \rho_k^{-1}\boldsymbol{\Delta}_k) - \boldsymbol{E}\|_F^2 + \rho_k^{-1}\|\boldsymbol{E}\|_1. \tag{20}$$

The solution of $\boldsymbol{E}$ can be computed in closed-form as

$$\boldsymbol{E}_{k+1} = \mathcal{S}_{\rho_k^{-1}}(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}_{k+1} - \rho_k^{-1}\boldsymbol{\Delta}_k), \tag{21}$$

where $\mathcal{S}_\tau(x) = \text{sign}(x) * \max(|x| - \tau, 0)$ is the soft-thresholding operator.

(4) Obtain the Lagrangian multipliers $(\boldsymbol{\Delta}_{k+1}, \boldsymbol{\delta}_{k+1})$ while fixing $(\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1}, \boldsymbol{E}_{k+1})$:

$$
\begin{aligned}
\boldsymbol{\Delta}_{k+1} &= \boldsymbol{\Delta}_k + \tau\rho_k(\boldsymbol{E}_{k+1} - \boldsymbol{X} + \boldsymbol{X}\boldsymbol{C}_{k+1}), \\
\boldsymbol{\delta}_{k+1} &= \boldsymbol{\delta}_k + \tau\rho_k(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}),
\end{aligned} \tag{22}
$$

where $\tau \in (0, \frac{\sqrt{5}+1}{2})$ is the dual step size and is usually set as $\tau = 1$.

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu\rho_k$, where $\mu > 1$.

5 Large Scale Subset Selection Via Woodbury Identity

The Woodbury Identity is

$$
(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}. \tag{23}
$$

We can also restrict that $\text{diag}(\boldsymbol{A}) = \boldsymbol{0}$ to avoid the samples to be self-represented. However, I want to mention that the proposed model solved by ADMM algorithm with three variables and does not have convergence results.

Then the model above can be

$$
\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1}. \tag{24}
$$

By introducing an auxiliary variable $\boldsymbol{C}$ into the optimization program, we can get

$$
\min_{\boldsymbol{A},\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \text{ s.t. } \boldsymbol{C} = \boldsymbol{A}, \tag{25}
$$

whose solution for $\boldsymbol{A}$ coincides with the solution of Eq. (43). By introducing two Lagrangian multipliers $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (43) can be written as

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) &= \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{C}\|_{p,1} \\
&+ \langle\boldsymbol{\Delta}, \boldsymbol{C} - \boldsymbol{A}\rangle + \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A}\|_F^2
\end{aligned} \tag{26}
$$

Denote by $(\boldsymbol{C}_k, \boldsymbol{A}_k)$ the optimization variables at iteration k, by $\boldsymbol{\Delta}_k$ the Lagrangian multipliers at iteration k, and by ρ_k the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing $(\boldsymbol{C}_k, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$
\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{A} - (\boldsymbol{C}_k + \rho_k^{-1}\boldsymbol{\Delta}_k)\|_F^2, \tag{27}
$$

which is equalivalently to solve the following problem

$$
\boldsymbol{A} = (\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho_k}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho_k}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k) \tag{28}
$$

Since the matrices $\boldsymbol{X}^\top\boldsymbol{X}$ is of $N \times N$ dimension. It is computational expensive when N is very large. By employing the Woodburry Identity mentioned above, we can have

$$
(\frac{\rho_k}{2}\boldsymbol{I} + \boldsymbol{X}^\top\boldsymbol{X})^{-1} = \frac{2}{\rho_k}\boldsymbol{I} - (\frac{2}{\rho_k})^2\boldsymbol{X}^\top(\boldsymbol{I} + \frac{2}{\rho_k}\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}. \tag{29}
$$

and transform this problem as

$$
\begin{aligned}
\boldsymbol{A} = &(\frac{2}{\rho_k}\boldsymbol{I} - (\frac{2}{\rho_k})^2\boldsymbol{X}^\top(\boldsymbol{I} + \frac{2}{\rho_k}\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}) \\
&* (\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho_k}{2}\boldsymbol{C}_k + \frac{1}{2}\boldsymbol{\Delta}_k)
\end{aligned} \tag{30}
$$

which will save a lot of computational costs.

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing $(\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$
\min_{\boldsymbol{C}} \frac{1}{2}\|(\boldsymbol{A}_{k+1} - \rho_k^{-1}\boldsymbol{\Delta}k) - \boldsymbol{C}\|_F^2 + \frac{\lambda}{\rho_k}\|\boldsymbol{C}\|_{p,1}. \tag{31}
$$

Since the $\ell_{p,1}$ norm is separable with respect to each row, we can write the above problem as

$$
\min_{\boldsymbol{C}} \sum_{i=1}^{M} \frac{1}{2}\|(\boldsymbol{A}_{k+1})_{i*} - \rho_k^{-1}(\boldsymbol{\Delta}_k)_{i*} - \boldsymbol{C}_{i*}\|_2^2 + \frac{\lambda}{\rho_k}\|\boldsymbol{C}_{i*}\|_p, \tag{32}
$$

where $\boldsymbol{F}_{i*}$ is the ith row of the matrix $\boldsymbol{F}$. Since this step is separable w.r.t. each row, we can employ parallel processing resources and reduce its computational time.

(3) Obtain the Lagrangian multipliers $(\boldsymbol{\Delta}_{k+1})$ while fixing $(\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1})$:

$$
\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho_k(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}). \tag{33}
$$

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu\rho_k$, where $\mu > 1$.

3

6 Robust Large Scale Subset Selection via Dissimilarity based Outlier Detection

We can also introduce a dissimiarlity based matrix D to replace the ℓ_p or $\ell_{2,1}$ norms to ensure robustness. This can also remore the additional term Z on modeling the outliers with the restriction of ℓ_1 norm. The matrix D should better be diagonal matrix. How to design the matrix D is another problem need to be solved.

Then the proposed model can be formulated as

$$\min_{A} \|(X - XA)D\|_F^2 + \lambda\|A\|_{p,1}. \qquad (34)$$

By introducing an auxiliary variable C into the optimization program, we can get

$$\min_{A,C} \|(X - XA)D\|_F^2 + \lambda\|C\|_{p,1} \text{ s.t. } C = A. \quad (35)$$

By introducing a Lagrangian multiplier Δ, the Lagrangian function of the Eq. (43) can be written as

$$\mathcal{L}(A, C, \Delta, \rho) = \|(X - XA)D\|_F^2 + \lambda\|C\|_{p,1}$$
$$+ \langle \Delta, C - A \rangle + \frac{\rho}{2}\|C - A\|_F^2 \qquad (36)$$

Denote by (A_k, C_k) the optimization variables at iteration k, by Δ_k the Lagrangian multiplier at iteration k, and by ρ_k the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain A_{k+1} by minimizing $\mathcal{L}$ with respect to A, while fixing (C_k, Δ_k). This is equivalent to solve the following problem:

$$\min_{A} \|(X - XA)D\|_F^2 + \frac{\rho}{2}\|A - (C_k - \rho_k^{-1}\Delta_k)\|_F^2, \qquad (37)$$

which is equalivalently to solve the following problem

$$X^\top XADD^\top + \frac{\rho_k}{2}A = X^\top XDD^\top + \frac{\rho_k}{2}(C_k - \rho_k^{-1}\Delta_k) \qquad (38)$$

Since the matrices $X^\top X$ and $D^\top D$ are positive semi-definite and positive definite, respectively. The above equation is a standard Sylvester equation which has a unique solution.

(2) Obtain C_{k+1} by minimizing $\mathcal{L}$ with respect to C, while fixing (A_{k+1}, Δ_k). This is equivalent to solve the following problem:

$$\min_{C} \frac{1}{2}\|(A_{k+1} + \rho_k^{-1}\Delta k) - C\|_F^2 + \frac{\lambda}{\rho_k}\|C\|_{p,1}. \quad (39)$$

Since the $\ell_{p,1}$ norm is separable with respect to each row, we can write the above problem as

$$\min_{C} \sum_{i=1}^{M} \frac{1}{2}\|(A_{k+1})_{i*} + \rho_k^{-1}(\Delta_k)_{i*} - C_{i*}\|_2^2 + \frac{\lambda}{\rho_k}\|C_{i*}\|_p, \qquad (40)$$

where F_{i*} is the ith row of the matrix F. Since this step is separable w.r.t. each row, we can employ parallel processing resources and reduce its computational time.

(3) Obtain the Lagrangian multipliers (Δ_{k+1}) while fixing (C_{k+1}, A_{k+1}):

$$\Delta_{k+1} = \Delta_k + \rho_k(C_{k+1} - A_{k+1}). \qquad (41)$$

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu\rho_k$, where $\mu > 1$.

7 Large Scale Subset Selection Via Row-Column Separation

We can also restrict that $\text{diag}(A) = \mathbf{0}$ to avoid the samples to be self-represented. However, I want to mention that the proposed model solved by ADMM algorithm with three variables and does not have convergence results.

Then the model above can be

$$\min_{A} \|X - XA\|_F^2 + \lambda\|A\|_{p,1} \quad \text{s.t.} \quad \text{diag}(A) = \mathbf{0}. \qquad (42)$$

By introducing an auxiliary variable C into the optimization program, we can get

$$\min_{A,C} \|X - XC\|_F^2 + \lambda\|A\|_{p,1}$$
$$\text{s.t. } C = A - \text{diag}(A), \qquad (43)$$

whose solution for $\boldsymbol{A}$ coincides with the solution of Eq. (43). By introducing two Lagrangian multipliers $\boldsymbol{\Delta}$, the Lagrangian function of the Eq. (43) can be written as

$$\mathcal{L}(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{\Delta}, \rho) = \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \lambda\|\boldsymbol{A}\|_{p,1}$$
$$+ \langle \boldsymbol{\Delta}, \boldsymbol{C} - (\boldsymbol{A} - \operatorname{diag}(\boldsymbol{A})) \rangle + \frac{\rho}{2}\|\boldsymbol{C} - (\boldsymbol{A} - \operatorname{diag}(\boldsymbol{A}))\|_F^2$$

$$(44)$$

Denote by $(\boldsymbol{C}_k, \boldsymbol{A}_k)$ the optimization variables at iteration k, by $\boldsymbol{\Delta}_k$ the Lagrangian multipliers at iteration k, and by ρ_k the penalty parameter at iteration k. Taking detivatives of $\mathcal{L}$ with respect to the variables and setting the derivatives to be zeros, we can alternatively update the variables as follows:

(1) Obtain $\boldsymbol{A}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{A}$, while fixing $(\boldsymbol{C}_k, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\boldsymbol{A}_{k+1} = \boldsymbol{J} - \operatorname{diag}(\boldsymbol{J}),$$
$$\boldsymbol{J} = \arg\min_{\boldsymbol{J}} \frac{1}{2}\|\boldsymbol{C}_k + \rho_k^{-1}\boldsymbol{\Delta}_k - \boldsymbol{J}\|_F^2 + \frac{\lambda}{\rho_k}\|\boldsymbol{J}\|_{p,1}.$$

$$(45)$$

(2) Obtain $\boldsymbol{C}_{k+1}$ by minimizing $\mathcal{L}$ with respect to $\boldsymbol{C}$, while fixing $(\boldsymbol{A}_{k+1}, \boldsymbol{\Delta}_k)$. This is equivalent to solve the following problem:

$$\min_{\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \frac{\rho_k}{2}\|\boldsymbol{C} - \boldsymbol{A}_{k+1} + \frac{1}{\rho_k}\boldsymbol{\Delta}_k\|_F^2 \quad (46)$$

This is a least squares regression problem which has a closed-form solution as

$$\boldsymbol{C}_{k+1} = (\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho_k}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top\boldsymbol{X} + \frac{\rho_k}{2}\boldsymbol{A}_{k+1} - \frac{1}{2}\boldsymbol{\Delta}_k).$$

$$(47)$$

(3) Obtain the Lagrangian multipliers $(\boldsymbol{\Delta}_{k+1})$ while fixing $(\boldsymbol{C}_{k+1}, \boldsymbol{A}_{k+1})$:

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho_k(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1}). \quad (48)$$

(5) Update the penalty parameter ρ as $\rho_{k+1} = \mu\rho_k$, where $\mu > 1$.