

Positive Collaborative Representation for Subspace Clustering

June 10, 2017

Abstract

1 Introduction

2 Motivation

Problem 1: Positive collaborative representation could

Problem 2: The original model need to process all the data samples, which is not scalable to large datasets.

Problem 3: Block grouping effects of the least square regression is able to group

Problem 4:

Problem 5:

3 LSR Model

The proposed robust, efficient, and scalable (RES) subset selection model can be formulated as follows:

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{X}\mathbf{A}\|_p^p + \lambda \|\mathbf{A}\|_F^2, \quad (1)$$

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

Then the model above can be

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{X}\mathbf{A}\|_p^p + \lambda \|\mathbf{A}\|_F^2. \quad (2)$$

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

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

By introducing a Lagrangian multiplier Δ , the Lagrangian function of the Eq. (15) can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{A}, \mathbf{E}, \Delta, \rho) = & \|\mathbf{E}\|_p^p + \lambda \|\mathbf{A}\|_F^2 \\ & + \frac{\rho}{2} \|\mathbf{E} - \mathbf{X} + \mathbf{X}\mathbf{A} + \rho^{-1} \Delta\|_F^2 \end{aligned} \quad (4)$$

Denote by $(\mathbf{A}_k, \mathbf{E}_k)$ the optimization variables at iteration k , and by Δ the Lagrangian multipliers 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 \mathbf{A}_{k+1} by minimizing \mathcal{L} with respect to \mathbf{A} , while fixing (\mathbf{E}_k, Δ_k) . This is equivalent to solve the following problem:

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

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

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

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

$$\min_{\mathbf{E}} \frac{\rho_k}{2} \|\mathbf{Q}_k - \mathbf{E}\|_F^2 + \|\mathbf{E}\|_p^p, \quad (7)$$

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

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

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

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

where

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

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

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

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

$$\Delta_{k+1} = \Delta_k + \tau\rho_k(\mathbf{E}_{k+1} - \mathbf{X} + \mathbf{X}\mathbf{A}_{k+1}), \quad (12)$$

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 $\text{diag}(\mathbf{A}) = \mathbf{0}$ 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

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

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

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{E}} \quad & \|\mathbf{E}\|_p^p + \lambda\|\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \mathbf{E} = \mathbf{X} - \mathbf{X}\mathbf{A}, \text{diag}(\mathbf{A}) = \mathbf{0}. \end{aligned} \quad (14)$$

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

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{C}, \mathbf{E}} \quad & \|\mathbf{E}\|_p^p + \lambda\|\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \mathbf{E} = \mathbf{X} - \mathbf{X}\mathbf{C}, \mathbf{C} = \mathbf{A} - \text{diag}(\mathbf{A}) \end{aligned} \quad (15)$$

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

Δ, δ , the Lagrangian function of the Eq. (15) can be written as

$$\begin{aligned} \mathcal{L}(\mathbf{A}, \mathbf{C}, \mathbf{E}, \Delta, \delta, \rho) = & \|\mathbf{E}\|_p^p + \lambda\|\mathbf{A}\|_F^2 \\ & + \frac{\rho}{2}\|\mathbf{E} - \mathbf{X} + \mathbf{X}\mathbf{C} + \rho^{-1}\Delta\|_F^2 \\ & + \frac{\rho}{2}\|\mathbf{C} - \mathbf{A} + \text{diag}(\mathbf{A}) + \rho^{-1}\delta\|_F^2 \end{aligned} \quad (16)$$

Denote by $(\mathbf{C}_k, \mathbf{A}_k, \mathbf{E}_k)$ the optimization variables at iteration k , by (Δ_k, δ_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 \mathbf{A}_{k+1} by minimizing \mathcal{L} with respect to \mathbf{A} , while fixing $(\mathbf{C}_k, \mathbf{E}_k, \Delta_k, \delta_k)$. This is equivalent to solve the following problem:

$$\begin{aligned} \mathbf{A}_{k+1} &= \mathbf{J} - \text{diag}(\mathbf{J}), \\ \mathbf{J} &= (\rho_k + 2\lambda)^{-1}(\rho_k\mathbf{C}_k + \delta_k) \end{aligned} \quad (17)$$

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

$$\begin{aligned} \mathbf{C}_{k+1} = \arg \min_{\mathbf{C}} \quad & \frac{\rho_k}{2}\|\mathbf{E}_k - \mathbf{X} + \mathbf{X}\mathbf{C} + \rho_k^{-1}\Delta_k\|_F^2 \\ & + \frac{\rho_k}{2}\|\mathbf{C} - \mathbf{A}_{k+1} + \rho_k^{-1}\delta_k\|_F^2 \end{aligned} \quad (18)$$

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

$$\begin{aligned} \mathbf{C}_{k+1} &= \mathbf{H} - \text{diag}(\mathbf{H}), \\ (\mathbf{X}^\top\mathbf{X} + \mathbf{I})\mathbf{H} &= \mathbf{X}^\top\mathbf{P}_k + \mathbf{Q}_k, \end{aligned} \quad (19)$$

where $\mathbf{P}_k = \mathbf{X} - \mathbf{E}_k - \rho_k^{-1}\Delta_k$ and $\mathbf{Q}_k = \mathbf{A}_{k+1} - \rho_k^{-1}\delta_k$.

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

$$\min_{\mathbf{E}} \frac{1}{2}\|(\mathbf{X} - \mathbf{X}\mathbf{C}_{k+1} - \rho_k^{-1}\Delta_k) - \mathbf{E}\|_F^2 + \rho_k^{-1}\|\mathbf{E}\|_1. \quad (20)$$

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

$$\mathbf{E}_{k+1} = \mathcal{S}_{\rho_k^{-1}}(\mathbf{X} - \mathbf{X}\mathbf{C}_{k+1} - \rho_k^{-1}\Delta_k), \quad (21)$$

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

(4) Obtain the Lagrangian multipliers $(\Delta_{k+1}, \delta_{k+1})$ while fixing $(C_{k+1}, A_{k+1}, E_{k+1})$:

$$\begin{aligned}\Delta_{k+1} &= \Delta_k + \tau \rho_k (E_{k+1} - X + X C_{k+1}), \\ \delta_{k+1} &= \delta_k + \tau \rho_k (C_{k+1} - A_{k+1}),\end{aligned}\quad (22)$$

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

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \quad (23)$$

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

$$\min_A \|X - XA\|_F^2 + \lambda \|A\|_{p,1}. \quad (24)$$

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

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

whose solution for A coincides with the solution of Eq. (43). By introducing two Lagrangian multipliers Δ , the Lagrangian function of the Eq. (43) can be written as

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

Denote by (C_k, A_k) the optimization variables at iteration k , by Δ_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 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:

$$\min_A \|X - XA\|_F^2 + \frac{\rho}{2} \|A - (C_k + \rho_k^{-1} \Delta_k)\|_F^2, \quad (27)$$

which is equivalently to solve the following problem

$$A = (X^\top X + \frac{\rho_k}{2} I)^{-1} (X^\top X + \frac{\rho_k}{2} C_k + \frac{1}{2} \Delta_k) \quad (28)$$

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

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

and transform this problem as

$$\begin{aligned}A &= (\frac{2}{\rho_k} I - (\frac{2}{\rho_k})^2 X^\top (I + \frac{2}{\rho_k} X X^\top)^{-1} X) \\ &* (X^\top X + \frac{\rho_k}{2} C_k + \frac{1}{2} \Delta_k)\end{aligned}\quad (30)$$

which will save a lot of computational costs.

(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:

$$\min_C \frac{1}{2} \|(A_{k+1} - \rho_k^{-1} \Delta_k) - C\|_F^2 + \frac{\lambda}{\rho_k} \|C\|_{p,1}. \quad (31)$$

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

$$\min_C \sum_{i=1}^M \frac{1}{2} \|(A_{k+1})_{i*} - \rho_k^{-1} (\Delta_k)_{i*} - C_{i*}\|_2^2 + \frac{\lambda}{\rho_k} \|C_{i*}\|_p, \quad (32)$$

where F_{i*} is the i th 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}) :

$$\Delta_{k+1} = \Delta_k + \rho_k (C_{k+1} - A_{k+1}). \quad (33)$$

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

6 Robust Large Scale Subset Selection via Dissimilarity based Outlier Detection

We can also introduce a dissimilarity based matrix D to replace the ℓ_p or $\ell_{2,1}$ norms to ensure robustness. This can also remove 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

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

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

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

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

$$\begin{aligned} \mathcal{L}(\mathbf{A}, \mathbf{C}, \Delta, \rho) = & \|(\mathbf{X} - \mathbf{X}\mathbf{A})\mathbf{D}\|_F^2 + \lambda \|\mathbf{C}\|_{p,1} \\ & + \langle \Delta, \mathbf{C} - \mathbf{A} \rangle + \frac{\rho}{2} \|\mathbf{C} - \mathbf{A}\|_F^2 \end{aligned} \quad (36)$$

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

$$\min_{\mathbf{A}} \|(\mathbf{X} - \mathbf{X}\mathbf{A})\mathbf{D}\|_F^2 + \frac{\rho}{2} \|\mathbf{A} - (\mathbf{C}_k - \rho_k^{-1} \Delta_k)\|_F^2, \quad (37)$$

which is equivalently to solve the following problem

$$\mathbf{X}^\top \mathbf{X} \mathbf{A} \mathbf{D} \mathbf{D}^\top + \frac{\rho_k}{2} \mathbf{A} = \mathbf{X}^\top \mathbf{X} \mathbf{D} \mathbf{D}^\top + \frac{\rho_k}{2} (\mathbf{C}_k - \rho_k^{-1} \Delta_k) \quad (38)$$

Since the matrices $\mathbf{X}^\top \mathbf{X}$ and $\mathbf{D}^\top \mathbf{D}$ are positive semi-definite and positive definite, respectively. The above

equation is a standard Sylvester equation which has a unique solution.

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

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

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

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

where \mathbf{F}_{i*} is the i th row of the matrix \mathbf{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 $(\mathbf{C}_{k+1}, \mathbf{A}_{k+1})$:

$$\Delta_{k+1} = \Delta_k + \rho_k (\mathbf{C}_{k+1} - \mathbf{A}_{k+1}). \quad (41)$$

(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 $\text{diag}(\mathbf{A}) = \mathbf{0}$ 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

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

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

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{C}} & \|\mathbf{X} - \mathbf{X}\mathbf{C}\|_F^2 + \lambda \|\mathbf{A}\|_{p,1} \\ \text{s.t. } & \mathbf{C} = \mathbf{A} - \text{diag}(\mathbf{A}), \end{aligned} \quad (43)$$

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

$$\begin{aligned}\mathcal{L}(\mathbf{A}, \mathbf{C}, \mathbf{\Delta}, \rho) &= \|\mathbf{X} - \mathbf{XC}\|_F^2 + \lambda \|\mathbf{A}\|_{p,1} \\ &+ \langle \mathbf{\Delta}, \mathbf{C} - (\mathbf{A} - \text{diag}(\mathbf{A})) \rangle + \frac{\rho}{2} \|\mathbf{C} - (\mathbf{A} - \text{diag}(\mathbf{A}))\|_F^2\end{aligned}\quad (44)$$

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

$$\begin{aligned}\mathbf{A}_{k+1} &= \mathbf{J} - \text{diag}(\mathbf{J}), \\ \mathbf{J} &= \arg \min_{\mathbf{J}} \frac{1}{2} \|\mathbf{C}_k + \rho_k^{-1} \mathbf{\Delta}_k - \mathbf{J}\|_F^2 + \frac{\lambda}{\rho_k} \|\mathbf{J}\|_{p,1}.\end{aligned}\quad (45)$$

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

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

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

$$\mathbf{C}_{k+1} = (\mathbf{X}^\top \mathbf{X} + \frac{\rho_k}{2} \mathbf{I})^{-1} (\mathbf{X}^\top \mathbf{X} + \frac{\rho_k}{2} \mathbf{A}_{k+1} - \frac{1}{2} \mathbf{\Delta}_k). \quad (47)$$

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

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

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