

Supplementary Material: Jointly Learning Non-negative Projection and Dictionary with Discriminative Graph Constraints for Classification

Weiyang Liu¹, Zhiding Yu², Yandong Wen², Rongmei Lin³, Meng Yang⁴
¹Peking University, ²Carnegie Mellon University
³Emory University, ⁴Shenzhen University

Summary

The JNPDL model is well motivated by the current drawbacks of dictionary learning approaches, while each constraints are also well designed (the novel discriminative graph constraints are proposed, and all constrains are designed to be easily optimized). Aiming to bridge the gap between features and dictionary, we do think the proposed idea of learning a projection for features jointly with the dictionary is worth noticing. Consider that the given training data is usually not naturally discriminative, yhus the discrimination of the learned dictionary will be limited by the training data. Jointly learning the discriminative feature projection and discriminative dictionary together helps to improve each other. The key of JNPDL is to learn a discriminative projection (non-negativity is one way to discrimination) that can better work with the dictionary.

In the supplementary material, we give the detailed optimization framework for the JNPDL model in **Appendix A**, and then provide a detailed proof in **Appendix B** to show the convergence of the updating rules for P and M is theoretically guaranteed. In other words, the proposed multiplicative updating algorithm for the non-negative projection P in the paper is proved to be convergent.

Appendix A: Optimization Framework

We adopt a standard iterative learning framework to jointly learn the sparse representation X , the non-negative projection matrix P , the intermediate non-negative basis matrix M and the dictionary D . The proposed algorithm is shown in Algorithm 1. The non-negative projection learning converges as we prove in **Appendix B**.

Non-negative projection learning

To learn the non-negative projection, we optimize P, M with D, X fixed. Thus the JNPDL model is rewritten as

$$\begin{aligned} \min_{P \geq 0, M \geq 0} \{ & \|PY - DX\|_F^2 + \sum_{i=1}^K \|PY_i - D_i X_i\|_F^2 \\ & + \alpha_1 \|Y - MPY\|_F^2 + \alpha_1 \beta \cdot \text{Tr}(\hat{P}Y L_p Y^T \hat{P}^T) \\ & + \alpha_1 \beta \cdot \text{Tr}(\tilde{P}Y L_p^p Y^T \tilde{P}^T) + \alpha_1 \|M - P^T\|_F^2 \} \end{aligned} \quad (1)$$

which is essentially a projective non-negative matrix factorization problem [6, 2]. We use the multiplicative iterative solution [6, 2, 4] to solve Eq. (1). Specifically, we transform it into tractable sub-problems and optimize M and P by a multiplicative non-negative iterative procedure.

Because M is the basis matrix, it is necessary to require each column m_i to have unit l_2 norm, i.e., $\|m_i\| = 1$. This extra constraint makes the optimization more complicated, so we compensate the norms of the basis matrix into the coefficient matrix as in [4] and replace $\alpha_1 \beta \text{Tr}(\hat{P}Y L_p Y^T \hat{P}^T) + \alpha_1 \beta \text{Tr}(\tilde{P}Y L_p^p Y^T \tilde{P}^T)$ with

$$\alpha_1 \beta \cdot (\text{Tr}(\hat{Q} \hat{P} Y L_p Y^T \hat{P}^T \hat{Q}^T) + \text{Tr}(\tilde{Q} \tilde{P} Y L_p^p Y^T \tilde{P}^T \tilde{Q}^T)) \quad (2)$$

where \hat{Q} equals $\text{diag}\{\|m_1\|, \dots, \|m_q\|\}$ and \tilde{Q} equals $\text{diag}\{\|m_{q+1}\|, \dots, \|m_s\|\}$.

On optimizing M with P, D, X fixed. We can further rewrite Eq. (18) as $\text{Tr}(MG_m M^T)$ where

$$\begin{aligned} G_m &= G_{m+} - G_{m-} \\ &= \begin{bmatrix} \hat{P}Y(\alpha_1\beta B_p)Y^T \hat{P}^T & 0 \\ 0 & \tilde{P}Y(\alpha_1\beta B_p^p)Y^T \tilde{L}^T \end{bmatrix} \odot I \\ &\quad - \begin{bmatrix} \hat{P}Y(\alpha_1\beta W_p)Y^T \hat{P}^T & 0 \\ 0 & \tilde{P}Y(\alpha_1\beta W_p^p)Y^T \tilde{L}^T \end{bmatrix} \odot I \end{aligned} \quad (3)$$

where \odot denotes the element-wise matrix multiplication, and I is an identity matrix. Then we put the non-negative constraints into the objective function with respect to M , and define ψ_{ij} as the Lagrange multiplier for $M \geq 0$. With $\Psi = [\psi_{ij}]$, the Lagrange $\mathcal{L}(M)$ is defined as

$$\begin{aligned} \mathcal{L}(M) &= \|PY - DX\|_F^2 + \sum \|PY_i - D_i X_i^i\|_F^2 + \\ &\quad \alpha_1 \|Y - MPY\|_F^2 + \text{Tr}(MG_m M^T) \\ &\quad + \alpha_1 \|M - P^T\|_F^2 + \text{Tr}(\Psi M^T) \end{aligned} \quad (4)$$

Thus the partial derivative of \mathcal{L} with respect to M is

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial M} &= -2\alpha_1 YY^T P^T + 2\alpha_1 MPYY^T P^T \\ &\quad + 2MG_m + 2\alpha_1 M - 2\alpha_1 P^T + \Psi \end{aligned} \quad (5)$$

According to the Karush-Kuhn-Tucker (KKT) condition ($\psi_{ij} M_{ij} = 0$) and $\frac{\partial \mathcal{L}}{\partial M} = 0$, we can obtain the update rule:

$$M_{ij}^{(t+1)} = M_{ij}^{(t)} \frac{(\alpha_1 YY^T P^T + MG_m + \alpha_1 P^T)_{ij}}{(\alpha_1 MPYY^T P^T + MG_m + \alpha_1 M)_{ij}}. \quad (6)$$

On optimizing P with M, D, X fixed. After updating M , we normalize the column vectors of M and multiply the norm to the projective matrix P , namely $P_i \leftarrow P_i \times \|m_i\|_2$, $m_i \leftarrow m_i / \|m_i\|_2, \forall i$. By using the normalized M , we simplify Eq. (18) as $\alpha_1 \beta \text{Tr}(\hat{P}Y L_p Y^T \hat{P}^T) + \alpha_1 \beta \text{Tr}(\tilde{P}Y L_p^p Y^T \tilde{P}^T)$. Thus the Lagrange $\mathcal{L}(P)$ is

$$\begin{aligned} \mathcal{L}(P) &= \|PY - DX\|_F^2 + \sum \|PY_i - D_i X_i^i\|_F^2 + \\ &\quad \alpha_1 \beta \text{Tr}(\hat{P}Y L_p Y^T \hat{P}^T) + \alpha_1 \beta \text{Tr}(\tilde{P}Y L_p^p Y^T \tilde{P}^T) + \\ &\quad \alpha_1 \|Y - MPY\|_F^2 + \alpha_1 \|M - P^T\|_F^2 + \text{Tr}(\Phi P^T) \end{aligned} \quad (7)$$

where ϕ is the Lagrange multiplier for constraint $P_{ij} \geq 0$ and $\Phi = [\phi_{ij}]$. After setting $\frac{\partial \mathcal{L}}{\partial P} = 0$ and applying KKT condition ($\phi_{ij} P_{ij} = 0$), we obtain the update rule for P :

$$P_{ij}^{(t+1)} = P_{ij}^{(t)} \frac{\left(\begin{array}{c} DXY^T + \sum D_i X_i^i Y_i^T + \alpha_1 M^T YY^T \\ + \alpha_1 M^T + \alpha_1 \beta \left[\begin{array}{c} \hat{P}^t Y W_p Y^T \\ \tilde{P}^t Y W_p^p Y^T \end{array} \right] \end{array} \right)_{ij}}{\left(\begin{array}{c} P^t YY^T + \sum P^t Y_i Y_i^T + \alpha_1 P^t + \\ \alpha_1 M^T M P^t YY^T + \alpha_1 \beta \left[\begin{array}{c} \hat{P}^t Y B_p Y^T \\ \tilde{P}^t Y B_p^p Y^T \end{array} \right] \end{array} \right)_{ij}}. \quad (8)$$

Now both Eq. (6) and Eq. (41) are non-negative update. We prove that the convergence of the updating rule for P and M can be guaranteed. Detailed proof can refer to the supplementary material (Appendix A).

Discriminative dictionary learning

The discriminative dictionary learning is optimized using a standard iterative optimization procedure that is widely adopted in sparse coding.

On optimizing X with D, P, M fixed. With D, P fixed, the optimization of the JNPDL model becomes

$$\begin{aligned} \min_X \{ & \|PY - DX\|_F^2 + \sum_{i=1}^K \|PY_i - D_i X_i^i\|_F^2 + \alpha_3 \|X\|_1 \\ & + \sum_{i=1}^K \sum_{j=1, j \neq i}^K \|D_j X_i^j\|_F^2 + \alpha_2 \text{Tr}(X^T (L') X) \} \end{aligned} \quad (9)$$

Algorithm 1 Training Procedure of JNPDL

Input: Training samples $\mathbf{Y} = \|\mathbf{Y}_1, \dots, \mathbf{Y}_N\|$, intrinsic graph $\mathbf{W}_c, \mathbf{W}_p$, penalty graph $\mathbf{W}_c^p, \mathbf{W}_p^p$, parameters $\alpha_1, \alpha_2, \alpha_3, \beta$, iteration number T .

Output: Non-negative projection matrix \mathbf{P} , dictionary \mathbf{D} , coding coefficient matrix \mathbf{X} .

Step1: Initialization

- 1: $t = 1$.
- 2: Randomly initialize columns in $\mathbf{D}^0, \mathbf{M}^0$ with unit l_2 norm.
- 3: Initialize $\mathbf{x}_{i,1 \leq i \leq N}$ with $((\mathbf{D}^0)^T(\mathbf{D}^0) + \lambda_2 \mathbf{I})^{-1}(\mathbf{D}^0)^T \mathbf{y}_i$ where \mathbf{y}_i is the i th training sample (regardless of label).

Step2: Search local optima

- 4: **while** not convergence or $t < T$ **do**
- 5: Solve $\mathbf{P}^t, \mathbf{M}^t$ iteratively with fixed \mathbf{D}^{t-1} and \mathbf{X}^{t-1} via Eq. (1).
- 6: Solve \mathbf{X}^t with fixed $\mathbf{M}^t, \mathbf{D}^{t-1}$ and \mathbf{P}^t via Eq. (10).
- 7: Solve \mathbf{D}^t with fixed $\mathbf{M}^t, \mathbf{P}^t$ and \mathbf{X}^t via Eq. (12).
- 8: $t \leftarrow t + 1$.
- 9: **end while**

Step3: Output

- 10: Output $\mathbf{P} = \mathbf{P}^t, \mathbf{D} = \mathbf{D}^t$ and $\mathbf{X} = \mathbf{X}^t$.
-

where $\mathbf{L}' = \mathbf{L}_c - \mathbf{L}_c^p$. Eq. (9) can be solved using feature sign search algorithm [1] after certain formulation based on [7, 3]. We optimize \mathbf{X} class by class. Following [1], we update \mathbf{X}_i one by one in the i th class. We define $\mathbf{x}_{i,j}$ as the coding coefficients of the j th sample in the i th class and reformulated the problem as

$$\begin{aligned} \min_{\mathbf{X}} \{ & \|\mathbf{P}\mathbf{Y}_i - \mathbf{D}\mathbf{x}_{i,j}\|_F^2 + \|\mathbf{P}\mathbf{Y}_i - \mathbf{D}_i\mathbf{x}_{i,j}^i\|_F^2 \\ & + \sum_{n=1, n \neq i}^K \|\mathbf{D}_n\mathbf{x}_{i,j}^n\|_F^2 + \alpha_2 Q(\mathbf{x}_{i,j}) + \alpha_3 \|\mathbf{x}_{i,j}\|_1 \end{aligned} \quad (10)$$

where $Q(\mathbf{x}_{i,j}) = \alpha_2(\mathbf{x}_{i,j}^T \mathbf{X}_i \mathbf{L}'_j + (\mathbf{X}_i \mathbf{L}'_j)^T \mathbf{x}_{i,j} - \mathbf{x}_{i,j}^T \mathbf{L}'_{jj})$ in which \mathbf{L}'_i is the i th column of \mathbf{L} , and \mathbf{L}'_{ii} is the entry in the i th row and i th column of \mathbf{L} . $\mathbf{x}_{i,j}$ can be solved via feature sign search algorithm as in [7, 3].

On optimizing \mathbf{D} with $\mathbf{P}, \mathbf{X}, \mathbf{M}$ fixed. By fixing \mathbf{P}, \mathbf{M} and \mathbf{X} , the JNPDL model is rewritten as

$$\begin{aligned} \min_{\mathbf{D}} \{ & \|\mathbf{P}\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \sum_{i=1}^K \|\mathbf{P}\mathbf{Y}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2 \\ & + \sum_{i=1}^K \sum_{j=1, j \neq i}^K \|\mathbf{D}_j \mathbf{X}_i^j\|_F^2 \end{aligned} \quad (11)$$

for which we update \mathbf{D} class by class sequentially. When we update \mathbf{D}_v , the sub-dictionaries $\mathbf{D}_i, i \neq v$ associated to the other classes will be fixed. Thus Eq. (11) can be further rewritten to

$$\begin{aligned} \min_{\mathbf{D}_i, i \in \{1, 2, \dots, K\}} \{ & \|\mathbf{P}\mathbf{Y}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2 + \|\mathbf{P}\mathbf{Y}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2 \\ & + \sum_{j=1, j \neq i}^K \|\mathbf{D}_j \mathbf{X}_i^j\|_F^2 \end{aligned} \quad (12)$$

which is essentially a quadratic programming problem and can be directly solved by the algorithm presented in [5] (update \mathbf{D}_i atom by atom). Note that each atom in the dictionary should have unit l_2 norm.

Appendix B: Proof of the convergence of updating rules for \mathbf{P} and \mathbf{M}

Proof. Before proving the convergence of the updating rules, we first introduce some necessary preliminaries.

Definition 1. Function $\mathcal{G}(\mathbf{A}, \mathbf{A}')$ is an auxiliary function for function $\mathcal{F}(\mathbf{A})$ if the conditions

$$\mathcal{G}(\mathbf{A}, \mathbf{A}') \geq \mathcal{F}(\mathbf{A}), \quad \mathcal{G}(\mathbf{A}, \mathbf{A}) = \mathcal{F}(\mathbf{A}) \quad (13)$$

are satisfied.

Lemma 1. If $\mathcal{G}(\mathbf{A}, \mathbf{A}')$ is an auxiliary function for $\mathcal{F}(\mathbf{A})$, then $\mathcal{F}(\mathbf{A})$ is non-increasing to \mathbf{A} under the update

$$\mathbf{A}^{t+1} = \arg \min_{\mathbf{A}} \mathcal{G}(\mathbf{A}, \mathbf{A}^t) \quad (14)$$

where t denotes the t th iteration.

Proof. From Eq.(14), we construct the following relation:

$$\mathcal{F}(\mathbf{A}^t) = \mathcal{G}(\mathbf{A}^t, \mathbf{A}^t) \geq \mathcal{G}(\mathbf{A}^{t+1}, \mathbf{A}^t). \quad (15)$$

Because $\mathcal{G}(\mathbf{A}, \mathbf{A}')$ is an auxiliary function for $\mathcal{F}(\mathbf{A})$, we can obtain the following inequality from Eq. (13):

$$\mathcal{G}(\mathbf{A}^{t+1}, \mathbf{A}^t) \geq \mathcal{F}(\mathbf{A}^{t+1}) \quad (16)$$

which leads to

$$\mathcal{F}(\mathbf{A}^{t+1}) \leq \mathcal{F}(\mathbf{A}^t). \quad (17)$$

Thus $\mathcal{F}(\mathbf{A})$ is non-increasing with respect to \mathbf{A} under the updating rule in Eq. (14). The lemma is proved \square

We first consider the scenario when \mathbf{P} is fixed. With \mathbf{P} fixed, we rewrite the optimization objective (Eq. (10) in the paper) as

$$\begin{aligned} \mathcal{F}(\mathbf{M}) = & \alpha_1 \|\mathbf{Y} - \mathbf{M}\mathbf{P}\mathbf{Y}\|_F^2 + \text{Tr}(\mathbf{M}\mathbf{G}_m\mathbf{M}^T) \\ & + \alpha_1 \|\mathbf{M} - \mathbf{P}^T\|_F^2 \end{aligned} \quad (18)$$

We denote \mathcal{F}_{ij} as the part of $\mathcal{F}(\mathbf{M})$ relevant to \mathbf{M}_{ij} , and then compute the first-order and the second-order derivative as follows:

$$\begin{aligned} \mathcal{F}'_{ij}(\mathbf{M}) = & \alpha_1 (-2\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T + 2\mathbf{M}\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T)_{ij} \\ & + (2\mathbf{M}\mathbf{G}_m)_{ij} + \alpha_1 (2\mathbf{M} - 2\mathbf{P}^T)_{ij} \end{aligned} \quad (19)$$

$$\mathcal{F}''_{ij}(\mathbf{M}) = 2\alpha_1 (\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T + \mathbf{G}_m + \mathbf{I})_{jj} \quad (20)$$

where \mathbf{I} denotes an identity matrix with matched size. We construct the function $\mathcal{G}(\mathbf{M}_{ij}, \mathbf{M}_{ij}^t)$ as

$$\begin{aligned} \mathcal{G}(\mathbf{M}_{ij}, \mathbf{M}_{ij}^t) = & \mathcal{F}_{ij}(\mathbf{M}_{ij}^t) + \mathcal{F}'_{ij}(\mathbf{M}_{ij}^t)(\mathbf{M}_{ij} - \mathbf{M}_{ij}^t) \\ & + \frac{\alpha_1 (\mathbf{M}^t\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T + \mathbf{M}^t\mathbf{G}_{m+} + \mathbf{M}^t)_{ij}}{\mathbf{M}_{ij}^t} (\mathbf{M}_{ij} - \mathbf{M}_{ij}^t)^2 \end{aligned} \quad (21)$$

Lemma 2. $\mathcal{G}(\mathbf{M}_{ij}, \mathbf{M}_{ij}^t)$ in Eq. (21) is an auxiliary function for the function $\mathcal{F}_{ij}(\mathbf{M})$.

Proof. Because it is easily obtained that $\mathcal{G}(\mathbf{M}_{ij}, \mathbf{M}_{ij}) = \mathcal{F}_{ij}(\mathbf{M}_{ij})$, we only need to prove that $\mathcal{G}(\mathbf{M}_{ij}, \mathbf{M}_{ij}^t) \geq \mathcal{F}_{ij}(\mathbf{M}_{ij})$. We first compute the Taylor series expansion of $\mathcal{F}_{ij}(\mathbf{M})$ as

$$\begin{aligned} \mathcal{F}_{ij}(\mathbf{M}_{ij}) = & \mathcal{F}_{ij}(\mathbf{M}_{ij}^t) + \mathcal{F}'_{ij}(\mathbf{M}_{ij}^t)(\mathbf{M}_{ij} - \mathbf{M}_{ij}^t) \\ & + \frac{1}{2} \mathcal{F}''_{ij}(\mathbf{M}_{ij}^t)(\mathbf{M}_{ij} - \mathbf{M}_{ij}^t)^2 \end{aligned} \quad (22)$$

Because the following inequalities are satisfied:

$$\begin{aligned} (\mathbf{M}^t\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T)_{ij} &= \sum_v (\mathbf{M}_{iv}^t(\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T)_{vj}) \\ &\geq \mathbf{M}_{ij}^t(\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T)_{jj}, \end{aligned} \quad (23)$$

$$\begin{aligned} (\mathbf{M}^t\mathbf{G}_{m+})_{ij} &= \sum_v (\mathbf{M}_{iv}^t(\mathbf{G}_{m+})_{vj}) \\ &\geq \mathbf{M}_{ij}^t(\mathbf{G}_m)_{jj}, \end{aligned} \quad (24)$$

$$\mathbf{M}_{ij}^t \geq \mathbf{M}_{ij}^t \mathbf{I}_{jj}, \quad (25)$$

we can let the following relation hold:

$$\begin{aligned} & \frac{\alpha_1 (\mathbf{M}^t\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T + \mathbf{M}^t\mathbf{G}_{m+} + \mathbf{M}^t)_{ij}}{\mathbf{M}_{ij}^t} \\ & \geq (\mathbf{P}\mathbf{Y}\mathbf{Y}^T\mathbf{P}^T + \mathbf{G}_m)_{jj}. \end{aligned} \quad (26)$$

Therefore, we can prove that $\mathcal{G}(\mathbf{M}_{ij}, \mathbf{M}_{ij}^t) \geq \mathcal{F}_{ij}(\mathbf{M}_{ij})$ holds. The lemma is proved. \square

Theorem 1. The updating rule for M can be obtained by minimizing the auxiliary function $\mathcal{G}(M_{ij}, M_{ij}^t)$.

Proof. We let the derivative of $\mathcal{G}(M_{ij}, M_{ij}^t)$ with respect to M_{ij} equal to zero, namely

$$\begin{aligned} & \frac{\partial \mathcal{G}(M_{ij}, M_{ij}^t)}{\partial M_{ij}} \\ &= \frac{2\alpha_1(M^t P Y Y^T P^T + M^t G_{m+} + M^t)_{ij}}{M_{ij}^t} (M_{ij} - M_{ij}^t) \\ & \quad + \mathcal{F}'_{ij}(M_{ij}^t). \\ &= 0 \end{aligned} \quad (27)$$

from which we can derive

$$M_{ij}^{t+1} = M_{ij}^t \frac{(\alpha_1 Y Y^T P^T + M^t G_{m+} + \alpha_1 P^T)_{ij}}{(\alpha_1 M^t P Y Y^T P^T + M^t G_{m+} + \alpha_1 M^t)_{ij}}. \quad (28)$$

which is identical to the updating rule that we use in the paper. Thus the lemma is proved \square

Then we consider the other scenario when M is fixed. After updating the matrix M via Eq. (28), we normalize the column vectors m_i of M and consequently convey the norm to the projective matrix P , namely

$$\begin{aligned} P_i &\leftarrow P_i \times \|m_i\| \\ m_i &\leftarrow m_i / \|m_i\| \end{aligned} \quad (29)$$

where P_i is the i th column vector of the projection matrix P . Considering Eq. (29) and the fixed M , we can rewrite the optimization objective (Eq. (10) in the paper) as

$$\begin{aligned} \mathcal{F}(P) &= \|PY - DX\|_F^2 + \sum \|PY_i - D_i X_i^i\|_F^2 + \\ & \quad \alpha_1 \beta \text{Tr}(\hat{P} Y L_p Y^T \hat{P}^T) + \alpha_1 \beta \text{Tr}(\tilde{P} Y L_p^p Y^T \tilde{P}^T) \\ & \quad + \alpha_1 \|Y - MPY\|_F^2 + \alpha_1 \|M - P^T\|_F^2 \end{aligned} \quad (30)$$

By denoting \mathcal{F}_{ij} as the part of $\mathcal{F}(P)$ relevant to P_{ij} , we have the following derivatives:

$$\begin{aligned} \mathcal{F}'_{ij}(P) &= 2(PY Y^T)_{ij} - 2(DX Y^T)_{ij} + 2(\sum PY_i Y_i^T)_{ij} \\ & \quad - 2\sum (D_i X_i^i Y_i^T)_{ij} - 2\alpha_1 (M^T Y Y^T)_{ij} + 2\alpha_1 (M^T M P Y Y^T)_{ij}, \\ & \quad + 2\alpha_1 \beta \begin{bmatrix} \hat{P} Y L_p Y^T \\ \tilde{P} Y L_p^p Y^T \end{bmatrix}_{ij} + \alpha_1 (2P - 2M^T)_{ij} \end{aligned} \quad (31)$$

$$\begin{aligned} \mathcal{F}''_{ij}(P) &= 2(Y Y^T)_{jj} + 2(\sum Y_i Y_i^T)_{jj} + 2\alpha_1 (M^T M)_{ii} (Y Y^T)_{jj} \\ & \quad + 2\alpha_1 \beta \begin{bmatrix} Y L_p Y^T \\ Y L_p^p Y^T \end{bmatrix}_{jj} + 2\alpha_1 I_{jj} \end{aligned} \quad (32)$$

The auxiliary function of $\mathcal{F}_{ij}(P)$ is designed as

$$\begin{aligned} \mathcal{G}(P_{ij}, P_{ij}^t) &= \mathcal{F}_{ij}(P_{ij}^t) + \mathcal{F}'_{ij}(P_{ij}^t)(P_{ij} - P_{ij}^t) \\ & \quad \left(P^t Y Y^T + \sum P^t Y_i Y_i^T + \alpha_1 (P^t) + \right. \\ & \quad \left. \alpha_1 M^T M P Y Y^T + \alpha_1 \beta \begin{bmatrix} \hat{P}^t Y B_p Y^T \\ \tilde{P}^t Y B_p^p Y^T \end{bmatrix} \right)_{ij} \\ & \quad + \frac{1}{P_{ij}^t} (P_{ij} - P_{ij}^t)^2 \end{aligned} \quad (33)$$

Lemma 3. $\mathcal{G}(P_{ij}, P_{ij}^t)$ in Eq. (33) is an auxiliary function for the function $\mathcal{F}_{ij}(P)$.

Proof. Because obviously $\mathcal{G}(P_{ij}, P_{ij}) = \mathcal{F}_{ij}(P_{ij})$, we only need to prove that $\mathcal{G}(P_{ij}, P_{ij}^t) = \mathcal{F}_{ij}(P_{ij})$. We first obtain the Taylor series expansion of $\mathcal{F}_{ij}(P)$ as

$$\begin{aligned} \mathcal{F}_{ij}(P_{ij}) &= \mathcal{F}_{ij}(P_{ij}^t) + \mathcal{F}'_{ij}(P_{ij}^t)(P_{ij} - P_{ij}^t) \\ & \quad + \frac{1}{2} \mathcal{F}''_{ij}(P_{ij}^t)(P_{ij} - P_{ij}^t)^2 \end{aligned} \quad (34)$$

Since the following relations hold:

$$\begin{aligned} (P^t Y Y^T)_{ij} &= \sum_v (P_{iv}^t (Y Y^T)_{vj}) \\ &\geq P_{ij}^t (Y Y^T)_{jj} \end{aligned} \quad (35)$$

$$\begin{aligned} \left(\sum_i \mathbf{P}^t \mathbf{Y}_i \mathbf{Y}_i^T\right)_{ij} &= \sum_v \left(\mathbf{P}_{iv}^t \left(\sum \mathbf{Y}_i \mathbf{Y}_i^T\right)_{vj}\right) \\ &\geq \mathbf{P}_{ij}^t \left(\sum \mathbf{Y}_i \mathbf{Y}_i^T\right)_{jj}, \end{aligned} \quad (36)$$

$$\begin{aligned} \left(\mathbf{M}^T \mathbf{M} \mathbf{P}^t \mathbf{Y} \mathbf{Y}^T\right)_{ij} &= \sum_v \left(\left(\mathbf{M}^T \mathbf{M} \mathbf{P}^t\right)_{iv} \left(\mathbf{Y} \mathbf{Y}^T\right)_{vj}\right) \\ &\geq \left(\mathbf{M}^T \mathbf{M} \mathbf{P}^t\right)_{ij} \left(\mathbf{Y} \mathbf{Y}^T\right)_{jj} \\ &= \sum_v \left(\left(\mathbf{M}^T \mathbf{M}\right)_{iv} \mathbf{P}_{vj}^t\right) \left(\mathbf{Y} \mathbf{Y}^T\right)_{jj}, \\ &\geq \mathbf{P}_{ij}^t \left(\mathbf{M}^T \mathbf{M}\right)_{ii} \left(\mathbf{Y} \mathbf{Y}^T\right)_{jj} \end{aligned} \quad (37)$$

$$\begin{aligned} \begin{bmatrix} \hat{\mathbf{P}}^t \mathbf{Y} \mathbf{B}_p \mathbf{Y}^T \\ \tilde{\mathbf{P}}^t \mathbf{Y} \mathbf{B}_p^p \mathbf{Y}^T \end{bmatrix}_{ij} &= \begin{cases} \sum_v \left(\hat{\mathbf{P}}_{iv}^t \left(\mathbf{Y} \mathbf{B}_p \mathbf{Y}^T\right)_{vj}\right), & \text{if } j \leq q \\ \sum_v \left(\tilde{\mathbf{P}}_{iv}^t \left(\mathbf{Y} \mathbf{B}_p^p \mathbf{Y}^T\right)_{vj}\right), & \text{otherwise} \end{cases} \\ &\geq \begin{cases} \hat{\mathbf{P}}_{ij}^t \left(\mathbf{Y} \mathbf{B}_p \mathbf{Y}^T\right)_{jj}, & \text{if } j \leq q \\ \tilde{\mathbf{P}}_{ij}^t \left(\mathbf{Y} \mathbf{B}_p^p \mathbf{Y}^T\right)_{jj}, & \text{otherwise} \end{cases} \\ &\geq \begin{cases} \hat{\mathbf{P}}_{ij}^t \left(\mathbf{Y} \mathbf{L}_p \mathbf{Y}^T\right)_{jj}, & \text{if } j \leq q \\ \tilde{\mathbf{P}}_{ij}^t \left(\mathbf{Y} \mathbf{L}_p^p \mathbf{Y}^T\right)_{jj}, & \text{otherwise} \end{cases} \\ &= \mathbf{P}_{ij}^t \begin{bmatrix} \mathbf{Y} \mathbf{L}_p \mathbf{Y}^T \\ \mathbf{Y} \mathbf{L}_p^p \mathbf{Y}^T \end{bmatrix}_{jj} \\ &\mathbf{P}_{ij}^t \geq \mathbf{P}_{ij}^t \mathbf{I}_{jj}, \end{aligned} \quad (38)$$

$$\mathbf{P}_{ij}^t \geq \mathbf{P}_{ij}^t \mathbf{I}_{jj}, \quad (39)$$

we can have $\mathcal{G}(\mathbf{P}_{ij}, \mathbf{P}_{ij}^t) \geq \mathcal{F}(\mathbf{P}_{ij})$. Therefore the lemma is proved. \square

Theorem 2. *The updating rule for \mathbf{P} can be obtained by minimizing the auxiliary function $\mathcal{G}(\mathbf{P}_{ij}, \mathbf{P}_{ij}^t)$.*

Proof. Let $(\partial \mathcal{G}(\mathbf{P}_{ij}, \mathbf{P}_{ij}^t)) / (\partial \mathbf{P}_{ij}) = 0$, and we have

$$\begin{aligned} &2 \left(\frac{\mathbf{P}^t \mathbf{Y} \mathbf{Y}^T + \sum \mathbf{P}^t \mathbf{Y}_i \mathbf{Y}_i^T + \alpha_1 (\mathbf{P}^t) + \alpha_1 \mathbf{M}^T \mathbf{M} \mathbf{P} \mathbf{Y} \mathbf{Y}^T + \alpha_1 \beta \begin{bmatrix} \hat{\mathbf{P}}^t \mathbf{Y} \mathbf{B}_p \mathbf{Y}^T \\ \tilde{\mathbf{P}}^t \mathbf{Y} \mathbf{B}_p^p \mathbf{Y}^T \end{bmatrix}_{ij}}{\mathbf{P}_{ij}^t} (\mathbf{P}_{ij} - \mathbf{P}_{ij}^t) \right. \\ &\left. + \mathcal{F}'_{ij}(\mathbf{P}_{ij}^t) \right) = 0 \end{aligned} \quad (40)$$

from which we can derive the updating rule for \mathbf{P}

$$\mathbf{P}_{ij}^{(t+1)} = \mathbf{P}_{ij}^{(t)} \frac{\left(\begin{array}{c} \mathbf{D} \mathbf{X} \mathbf{Y}^T + \sum \mathbf{D}_i \mathbf{X}_i^t \mathbf{Y}_i^T + \alpha_1 \mathbf{M}^T \mathbf{Y} \mathbf{Y}^T \\ + \alpha_1 \mathbf{M}^T + \alpha_1 \beta \begin{bmatrix} \hat{\mathbf{P}}^t \mathbf{Y} \mathbf{W}_p \mathbf{Y}^T \\ \tilde{\mathbf{P}}^t \mathbf{Y} \mathbf{W}_p^p \mathbf{Y}^T \end{bmatrix}_{ij} \end{array} \right)}{\left(\begin{array}{c} \mathbf{P}^t \mathbf{Y} \mathbf{Y}^T + \sum \mathbf{P}^t \mathbf{Y}_i \mathbf{Y}_i^T + \alpha_1 \mathbf{P}^t + \\ \alpha_1 \mathbf{M}^T \mathbf{M} \mathbf{P}^t \mathbf{Y} \mathbf{Y}^T + \alpha_1 \beta \begin{bmatrix} \hat{\mathbf{P}}^t \mathbf{Y} \mathbf{B}_p \mathbf{Y}^T \\ \tilde{\mathbf{P}}^t \mathbf{Y} \mathbf{B}_p^p \mathbf{Y}^T \end{bmatrix}_{ij} \end{array} \right)} \quad (41)$$

Thus the theorem is proved. \square

According to **Lemma 1**, **Theorem 1** and **Theorem 2**, we have proved that the convergence of the updating rules for \mathbf{P} and \mathbf{M} can be theoretically guaranteed. \square

References

- [1] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *NIPS*, 2006.
- [2] Xiaobai Liu, Shuicheng Yan, and Hai Jin. Projective nonnegative graph embedding. *IEEE T. IP*, 19(5):1126–1137, 2010.
- [3] Jiwen Lu, Gang Wang, Weihong Deng, and Pierre Moulin. Simultaneous feature and dictionary learning for image set based face recognition. In *ECCV*. 2014.
- [4] Changhu Wang, Zheng Song, Shuicheng Yan, Lei Zhang, and Hong-Jiang Zhang. Multiplicative nonnegative graph embedding. In *CVPR*, 2009.

- [5] Meng Yang, Lei Zhang, Jian Yang, and David Zhang. Metaface learning for sparse representation based face recognition. In *ICIP*, 2010.
- [6] Zhirong Yang and Erkki Oja. Linear and nonlinear projective nonnegative matrix factorization. *IEEE T. NN*, 21(5):734–749, 2010.
- [7] Miao Zheng, Jiajun Bu, Chun Chen, Can Wang, Lijun Zhang, Guang Qiu, and Deng Cai. Graph regularized sparse coding for image representation. *IEEE T. IP*, 20(5):1327–1336, 2011.