Supplementary Materials: Enhancing Incomplete Multi-view Learning via Adaptive Tensor Graph Completion

Heng Zhang[1], Xiaohong Chen[1*]

1. College of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
mail: {zhangheng37, lyandcxh}@nuaa.edu.cn

Solution of the Model

To efficiently solve our model (10) in the original paper, we need to introduce the auxiliary variable $\mathcal{Y}$ to split the interdependent term such that they can be solved independently. Thus, we can reformulate model (10) into the following equivalent form,

$$\begin{aligned}\min_{A,\{W_i\}_{i=1}^m,\mathcal{G}} \quad & \sum_{i=1}^{m}\left(\delta_i \left\|\left(A-W_i^\top X_i\right)P_i\right\|_F^2+\lambda\delta_i \operatorname{tr}\left(AL_{G_i}A^\top\right)\right)+\mu\|\mathcal{Y}\|_{\Phi,w,S_p}^p+\frac{\gamma}{2}\left\|P_\Omega(\mathcal{G})-P_\Omega(\mathcal{M})\right\|_F^2 \\ \text{s.t.} \quad & AA^\top=I,\quad G_i\geq 0,\quad G_i\mathbf{1}=\mathbf{1},\quad \mathcal{G}=\mathcal{Y},\quad \delta_i=\frac{n_i}{\sqrt{\left\|\left(A-W_i^\top X_i\right)P_i\right\|_F^2+\lambda\operatorname{tr}\left(AL_{G_i}A^\top\right)}}\end{aligned}. \tag{1}$$

Inspired by recent progress on alternating direction methods, we proposed an efficient algorithm based on the inexact augmented Lagrange multiplier (ALM) method to solve the Eq.(1), whose augmented Lagrange function is given by

$$\begin{aligned}L_\rho\left(A,\{W_i\}_{i=1}^m,\mathcal{G},\mathcal{Y},\mathcal{C}\right)=&\sum_{i=1}^{m}\delta_i\left[\left\|\left(A-W_i^\top X_i\right)P_i\right\|_F^2+\lambda\operatorname{tr}\left(AL_{G_i}A^\top\right)\right]+\mu\|\mathcal{Y}\|_{\Phi,w,S_p}^p+\frac{\gamma}{2}\left\|P_\Omega(\mathcal{G})-P_\Omega(\mathcal{M})\right\|_F^2 \\ &+\langle\mathcal{C},\mathcal{G}-\mathcal{Y}\rangle+\frac{\rho}{2}\|\mathcal{G}-\mathcal{Y}\|_F^2\end{aligned}. \tag{2}$$

where $\rho>0$ is the penalty parameter, $\mathcal{C}\in\mathbb{R}^{n\times m\times n}$ is Lagrange multipliers. This section uses superscripts to indicate the number of iterations, i.e. A^k or A^{k+1}, etc.

Updating A^{k+1} and $\{W_i^{k+1}\}_{i+1}^m$

To update A^{k+1} and $\{W_i^{k+1}\}_{i=1}^m$, we consider the following optimization problems,

$$A^{k+1},\left\{W_i^{k+1}\right\}_{i=1}^m=\underset{AA^\top=I,\{W_i\}_{i=1}^m}{\arg\min}\sum_{i=1}^{m}\left(\delta_i^k\left\|\left(A-W_i^\top X_i\right)P_i\right\|_F^2+\lambda\delta_i^k\operatorname{tr}\left(AL_{G_i^k}A^\top\right)\right). \tag{3}$$

After simple algebraic operations, the problem (3) can be transformed into the following eigenvalue problem,

$$\left(\sum_{i=1}^{m}\delta_i^k\left(P_i-P_iX_i^\top\left(X_iP_iX_i^\top\right)^{-1}X_iP_i+\lambda L_{G_i^k}\right)\right)A^\top=A^\top\Sigma, \tag{4}$$

where Σ is a diagonal matrix of eigenvalues. Then $A^\top$ can be obtained from the eigenvector corresponding to the first d largest eigenvalues, where d is the dimension of consistent low-dimensional representation (i.e. $A\in\mathbb{R}^{d\times n}$). With the A, we can obtain $W_i^{k+1}=\left(X_iP_iX_i^\top\right)^{-1}X_iP_iA^{k+1\top}$ for $i=1,2,\ldots,m$. Once A^{k+1} and $\{W_i^{k+1}\}_{i=1}^m$ are obtained, the $\{\delta_i\}_{i=1}^m$ can be updated according to the definition to obtain $\{\delta_i^{k+1}\}_{i=1}^m$.

Updating $\mathcal{G}^{k+1}$

To solve $\mathcal{G}^{k+1}$, we fix the other variable and solve the following optimization problem,

$$\mathcal{G}^{k+1}=\underset{G_i\geq 0,G_i\mathbf{1}=\mathbf{1}}{\arg\min}\lambda\sum_{i=1}^{m}\delta_i^{k+1}\operatorname{tr}\left(A^{k+1}L_{G_i}A^{k+1\top}\right)+\frac{\gamma}{2}\left\|P_\Omega(\mathcal{G})-P_\Omega(\mathcal{M})\right\|_F^2+\frac{\rho^k}{2}\left\|\mathcal{G}-\mathcal{Y}^k+\frac{\mathcal{C}^k}{\rho^k}\right\|_F^2. \tag{5}$$

The problem (5) is a least squares problem with constraints. Notice that each tube of the tensor of the problem (5) is independent, so it can be converted into $n\times m$ independent subproblems. Without loss of generality, we let $g\in\mathbb{R}^n$, $s\in\mathbb{R}^n$, $y\in\mathbb{R}^n$, $\omega\in\mathbb{R}^n$, $c\in\mathbb{R}^n$, and $t\in\mathbb{R}^n$ be a tube of tensor $\mathcal{G}$, $\mathcal{M}$, $\mathcal{Y}$, Ω, $\mathcal{C}$, and $\mathcal{T}$, respectively, where $\mathcal{T}_{(i,j,l)}^k=1/2\lambda\delta_l^{k+1}\|A_{(:,i)}^{k+1}-A_{(:,j)}^{k+1}\|_F^2$. Thus, the following optimization problem can be obtained,

$$g^{k+1} = \underset{g\geq 0, \mathbf{1}^\top g=1}{\arg\min} \frac{1}{\rho^k} t^{k\top} g + \frac{\gamma}{2\rho^k} \left\| P_\omega(g) - P_\omega(z) \right\|_F^2 + \frac{1}{2} \left\| g - y^k + \frac{c^k}{\rho^k} \right\|_F^2 . \quad (6)$$

The Lagrange function of Eq.(6) is

$$L(g,\alpha,\beta) = \frac{1}{\rho^k} t^{k\top} g + \frac{\gamma}{2\rho^k} \left\| P_\omega(g) - P_\omega(z) \right\|_F^2 + \frac{1}{2} \left\| g - y^k + \frac{c^k}{\rho^k} \right\|_F^2 - \alpha\left(\mathbf{1}^\top g - 1\right) - \beta^\top g . \quad (7)$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^n$ are multipliers. Taking the derivative w.r.t. g and setting it to zero, the following equation holds,

$$g - y^k + \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k}\left(P_\omega(g) - P_\omega(z)\right) + \frac{1}{\rho^k} t^k - \alpha\mathbf{1} - \beta = \mathbf{0} . \quad (8)$$

According to the KKT condition, i.e., $\mathbf{1}^\top g = 1$, we have

$$\alpha = \frac{1}{n}\left(1 - \mathbf{1}^\top y^k + \mathbf{1}^\top \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k}\mathbf{1}^\top\left(P_\omega(g) - P_\omega(z)\right) + \frac{1}{\rho^k}\mathbf{1}^\top t^k - \mathbf{1}^\top \beta\right). \quad (9)$$

Combining Eq.(8) and Eq.(9), we can obtain

$$g + \frac{\gamma}{\rho^k} P_\omega(g) = y^k - \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k} P_\omega(z) - \frac{1}{\rho^k} t^k - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\left(y^k - \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k} + P_\omega(z) - \frac{1}{\rho^k} t^k\right) + \frac{1}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}\mathbf{1}^\top\left(\frac{\gamma}{\rho^k} P_\omega(g) - \beta\right) + \beta \quad (10)$$

Further according to the complementary relaxation in the KKT condition, we have

$$g_i^{k+1} + \frac{\gamma}{\rho^k} P_\omega\left(g^{k+1}\right)_i = \left(u_i^k - \frac{1}{n}\mathbf{1}^\top u^k + \frac{1}{n} + v^k\right)_+ , \quad (11)$$

where $u^k = y^k - c^k/\rho^k + \gamma/\rho^k\, P_\omega(z) - t^k/\rho^k$, $v^k = 1/n\mathbf{1}^\top\left(\gamma/\rho^k\, P_\omega\left(g^{k+1}\right) - \beta\right)$. So we can get

$$g_i^{k+1} = \begin{cases} \left(u_i^k - \frac{1}{n}\mathbf{1}^\top u^k + \frac{1}{n} + v^k\right)_+ , & \text{if } i \in \omega, \\ \frac{\rho^k}{\rho^k + \gamma}\left(u_i^k - \frac{1}{n}\mathbf{1}^\top u^k + \frac{1}{n} + v^k\right)_+ , & \text{otherwise.} \end{cases} . \quad (12)$$

Since $\mathbf{1}^\top g^{k+1} = 1$, the value of v^k can be found by solving the roots of the following function,

$$f(v) = \sum_{i=1}^{n} g_i^{k+1} - 1. \quad (13)$$

The equation (13) can be solved by Newton's method. The first order derivative of $f(v)$ is

$$\nabla f(v) = \sum_{i=1}^{n} \nabla g_i^{k+1} , \quad (14)$$

where

$$\nabla g_i^{k+1} = \begin{cases} 1 & \text{if } i \in \omega, \\ \frac{\rho^k}{\rho^k + \gamma} & \text{otherwise.} \end{cases} \quad (15)$$

The solution of $\mathcal{G}$ is summarized in Algorithm 1.

Algorithm 1

Input: $\{X_i\}_{i=1}^m$, $\{P_i\}_{i=1}^m$, ω , $\mathcal{M}$, μ , λ .

Initialize: $u = y^k - c^k/\rho^k + \gamma/\rho^k\, P_\omega(z) - t^k/\rho^k$, $v^0 = 0$, $k = 0$

While $|f(v)| > 10^{-10}$

Evaluate the function $f(v^k)$ according to Eq.(13).

Compute the gradient of $f(v)$ at point v^k according to Eq.(14).

Update v by $v^{k+1} = v^k - f(v^k)/\nabla f(v^k)$.

$k \leftarrow k+1$

Endwhile

Update g via Eq.(12).

Output: $\mathcal{G}$

Updating $\mathcal{Y}^{k+1}$

Keeping all other variables fixed, $\mathcal{Y}^{k+1}$ can be updated by solving the following optimization problem,

$$\min_{\mathcal{Y}} \ \frac{\mu}{\rho^k}\|\mathcal{Y}\|_{\Phi,w,S_p}^p + \frac{1}{2}\left\|\mathcal{G}^{k+1} + \frac{\mathcal{C}^k}{\rho^k} - \mathcal{Y}\right\|_F^2 \tag{16}$$

To solve it, we introduce the Theorem 1, which is proved in the below section.

Theorem 1 : *Suppose $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $l=\min(n_1,n_2)$, $w_1 \ge w_2 \ge \cdots \ge w_l \ge 0$, $\Phi^\top\Phi = I \in \mathbb{R}^{r\times r}$. Given the model*

$$\min_{\mathcal{X}} \ \tau\|\mathcal{X}\|_{\Phi,w,S_p}^p + \frac{1}{2}\|\mathcal{X}-\mathcal{A}\|_F^2, \tag{17}$$

then the optimal solution to the model (17) is

$$\mathcal{X}^* = \left(\mathrm{fold}\left(\mathrm{unfold}\left(\mathcal{S}_{\tau,w,p}(\mathcal{A}_\Phi)\right);\mathrm{unfold}\left(\mathcal{A}_{\Phi^c}\right)\right)\right)_{\bar{\Phi}^\top}, \tag{18}$$

where $\mathcal{S}_{\tau,w,p}\left(\mathcal{A}_\Phi\right)$ is a tensor, which the i-th frontal slice is $S_{\tau,w,p}\left(A_\Phi^{(i)}\right)$.

Therefore, the optimal solution for Eq.(16) is

$$\mathcal{Y}^{k+1} = \left(\mathrm{fold}\left(\mathrm{unfold}\left(\mathcal{S}_{\frac{\mu}{\rho},w,p}(\mathcal{B}_\Phi^k)\right);\mathrm{unfold}\left(\mathcal{B}_{\Phi^c}^k\right)\right)\right)_{\bar{\Phi}^\top} \tag{19}$$

where $\mathcal{B}^k = \mathcal{G}^{k+1} + \mathcal{C}^k/\rho^k$, $\mathcal{S}_{\mu/\rho,w,p}(\mathcal{B}_\Phi^k)$ is as shown in the proof of Theorem 1 in the below section. Based on the description above, the pseudo-code is summarized in Algorithm 2.

Algorithm 2

Input: $\{X_i\}_{i=1}^m$, $\{P_i\}_{i=1}^m$, w, $\mathcal{M}$, λ, μ, γ

Initialize: $k=0$, $\eta=1.1$, $\rho^0=10^{-4}$, $\{\delta_i^0 = 1/m\}_{i=1}^m$, $\mathcal{C}^0=0$, $\mathcal{G}^0=\mathcal{Y}^0=\mathcal{M}$.

While not converged

Update A^{k+1} via Eq.(4).

Update W_i^{k+1} via $(X_iP_iX_i^\top)^{-1}X_iP_iA^{k+1\top}$ for all i.

Update δ_i^{k+1} via definition for all i.

Update $\mathcal{G}^{k+1}$ via Algorithm 1.

Update $\mathcal{Y}^{k+1}$ via Eq.(19).

Update $\mathcal{C}^{k+1}$ via $\mathcal{C}^{k+1} = \mathcal{C}^k + \rho^k\left(\mathcal{G}^{k+1} - \mathcal{Y}^{k+1}\right)$.

Update ρ^{k+1} via $\rho^{k+1} = \eta\rho^k$.

$k \leftarrow k+1$

Endwhile

Output: A, $\{W_i\}_{i=1}^m$

Proof of the Theorem 1

Before proving Theorem 1, we first give Lemma 1 and Lemma 2.

Lemma 1 [1]*: For the optimization problem*

$$\min_{x>0} \ \frac{1}{2}(x-\sigma)^2 + wx^p \tag{20}$$

with the given p and w, there is a specific threshold

$$\tau_p(w) = \left(2w(1-p)\right)^{\frac{1}{2-p}} + wp\left(2w(1-p)\right)^{\frac{p-1}{2-p}}. \tag{21}$$

We have the following conclusion:

1) *When $\sigma \le \tau_p(w)$, the optimal solution $x_p(\sigma,w)$ of Eq.(20) is 0.*
2) *When $\sigma > \tau_p(w)$, the optimal solution is $x_p(\sigma,w) = \mathrm{sign}(\sigma)S_p(\sigma,w)$, where $S_p(\sigma,w)$ can be obtained by solving $S_p(\sigma,w) - \sigma + wp(S_p(\sigma,w))^{p-1} = 0$.*

Lemma 2 [1]*: Let $Y = U_Y\Sigma_Y V_Y^\top$ be the SVD of $Y\in\mathbb{R}^{m\times n}, \tau>0$, $l=\min(m,n)$, $w_1 \ge w_2 \ge \cdots \ge w_l \ge 0$, an optimal global solution of the following problem,*

$$\min_X \ \tau\|X\|_{w,S_p}^p + \frac{1}{2}\|X-Y\|_F^2, \tag{22}$$

is

$$S_{\tau,w,p}\left(Y\right)=U_Y P_{\tau,w,p}\left(Y\right)V_Y^{\top}, \tag{23}$$

where $P_{\tau,w,p}(Y)=\mathrm{diag}(\gamma_1,\gamma_2,\ldots,\gamma_l)$ *and* $\gamma_i=x_p(\sigma_i(Y),\tau w_i)$ *which can be obtained by Lemma 1.*

Let's give the definition of $\|X\|_{w,S_p}^p$,

$$\|X\|_{w,S_p}=\left(\sum_{i=1}^{l}w_i\sigma_i\left(X\right)^p\right)^{\frac{1}{p}}. \tag{24}$$

The proof of Theorem 1 is given as follows. For a given semi-orthogonal matrix $\Phi\in\mathbb{R}^{n\times r}$, there exists a semi-orthogonal matrix $\Phi^c\in\mathbb{R}^{n\times(n-r)}$ satisfying $\bar{\Phi}^{\top}\bar{\Phi}=I$, where $\bar{\Phi}=[\Phi,\Phi^c]\in\mathbb{R}^{n\times n}$. According to the definition, we have

$$\begin{aligned}\mathcal{X}^* &=\arg\min_{\mathcal{X}}\quad \tau\|\mathcal{X}\|_{\Phi,w,S_p}^p+\frac{1}{2}\|\mathcal{X}-\mathcal{A}\|_F^2\\ &=\arg\min_{\mathcal{X}}\quad \tau\sum_{i=1}^{r}\left\|X_{\Phi}^{(i)}\right\|_{w,S_p}^p+\frac{1}{2}\left\|\mathcal{X}_{\bar{\Phi}}-\mathcal{A}_{\bar{\Phi}}\right\|_F^2\\ &=\arg\min_{\mathcal{X}}\quad \tau\sum_{i=1}^{r}\left\|X_{\Phi}^{(i)}\right\|_{w,S_p}^p+\frac{1}{2}\sum_{i=1}^{r}\left\|X_{\Phi}^{(i)}-A_{\Phi}^{(i)}\right\|_F^2+\frac{1}{2}\sum_{j=1}^{n_3-r}\left\|X_{\Phi^c}^{(j)}-A_{\Phi^c}^{(j)}\right\|_F^2\end{aligned}. \tag{25}$$

In Eq.(25), each variable $X_{\Phi}^{(i)}$ and $X_{\Phi^c}^{(j)}$ is independent. Thus, it can be divided into n_3 independent subproblems, i.e.,

$$\min_{\mathcal{X}}\quad \tau\left\|X_{\Phi}^{(i)}\right\|_{w,S_p}^p+\frac{1}{2}\left\|X_{\Phi}^{(i)}-A_{\Phi}^{(i)}\right\|_F^2,\qquad i=1,2,\cdots,r \tag{26}$$

and

$$\min_{\mathcal{X}}\quad \frac{1}{2}\left\|X_{\Phi^c}^{(j)}-A_{\Phi^c}^{(j)}\right\|_F^2,\qquad j=1,2,\cdots,n_3-r. \tag{27}$$

According to Lemma 2, the optimal solution of Eq.(26) and Eq.(27) are $X_{\Phi}^{(i)*}=U_{A_{\Phi}^{(i)}}P_{\tau,w,p}(A_{\Phi}^{(i)})V_{A_{\Phi}^{(i)}}^{\top}$ and $X_{\Phi^c}^{(j)*}=A_{\Phi^c}^{(j)}$, respectively. Therefore, we obtain the optimal solution

$$\begin{aligned}\mathcal{X}_{\bar{\Phi}}^* &=\mathrm{fold}\left(X_{\Phi}^{(1)*};X_{\Phi}^{(2)*};\cdots;X_{\Phi}^{(r)*};X_{\Phi^c}^{(1)*};X_{\Phi^c}^{(2)*};\cdots;X_{\Phi^c}^{(n_3-r)*}\right)\\ &=\mathrm{fold}\left(S_{\tau,w,p}\left(A_{\Phi}^{(1)}\right);\cdots;S_{\tau,w,p}\left(A_{\Phi}^{(r)}\right);A_{\Phi^c}^{(1)};\cdots;A_{\Phi^c}^{(n_3-r)}\right)\\ &=\mathrm{fold}\left(\mathrm{unfold}\left(\mathcal{S}_{\tau,w,p}\left(\mathcal{A}_{\Phi}\right)\right);\mathrm{unfold}\left(\mathcal{A}_{\Phi^c}\right)\right)\end{aligned}. \tag{28}$$

Then, the optimal solution of Eq.(17) is $\mathcal{X}^*=(\mathcal{X}_{\bar{\Phi}}^*)_{\bar{\Phi}^{\top}}$.

References

[1] Y. Xie, S. Gu, Y. Liu *et al.*, "Weighted schatten p -norm minimization for image denoising and background subtraction," *IEEE transactions on image processing*, vol. 25, no. 10, pp. 4842-4857, 2016.