Supplementary Material of Consistent and Specific Multi-view Subspace Clustering

Shirui Luo[1,2], Changqing Zhang[3], Wei Zhang[1], Xiaochun Cao[1,2*]
[1]SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences
[3]School of Computer Science and Technology, Tianjin University, Tianjin, China
luoshirui@iie.ac.cn, zhangchangqing@tju.edu.cn, {wzhang, caoxiaochun}@iie.ac.cn

Optimization

In this section, we present the detailed optimization procedure of Section **The Proposed Approach** in full version.

The objective function of the proposed method is formulated as

$$\min_{\substack{\mathbf{C},\mathbf{D}^{(1)},\cdots,\mathbf{D}^{(V)},\\ \mathbf{E}^{(1)},\cdots,\mathbf{E}^{(V)}}} \sum_{v=1}^{V}\|\mathbf{E}^{(v)}\|_F^2+\lambda_C\|\mathbf{C}\|_*+\lambda_D\sum_{v=1}^{V}\|\mathbf{D}^{(v)}\|_2^2 \quad s.t.\ \mathbf{X}^{(v)}=\mathbf{X}^{(v)}(\mathbf{C}+\mathbf{D}^{(v)})+\mathbf{E}^{(v)}, \tag{1}$$

which simultaneously learns the consistent and specific representations from multiple views. However, directly finding the optimal solution of the problem (1) is extremely difficult. Thus, we propose a convex relaxation and develop an alternating optimization algorithm to jointly recover the corresponding data representations. Firstly, we introduce a variable $\mathbf{K}\in\mathbb{R}^{N\times N}$ as a surrogate of $\mathbf{C}$ in the nuclear norm:

$$\begin{aligned}\min_{\substack{\mathbf{C},\mathbf{D}^{(1)},\cdots,\mathbf{D}^{(V)},\\ \mathbf{E}^{(1)},\cdots,\mathbf{E}^{(V)}}} &\sum_{v=1}^{V}\|\mathbf{E}^{(v)}\|_F^2+\lambda_C\|\mathbf{K}\|_*+\lambda_D\sum_{v=1}^{V}\|\mathbf{D}^{(v)}\|_2^2\\ s.t.\ &\mathbf{X}^{(v)}=\mathbf{X}^{(v)}(\mathbf{C}+\mathbf{D}^{(v)})+\mathbf{E}^{(v)},\\ &\mathbf{C}=\mathbf{K}.\end{aligned} \tag{2}$$

This problem can be solved by the following Augmented Lagrange Multiplier (ALM) (Lin, Chen, and Ma 2010) method, which minimizes the augmented Lagrange function of the form:

$$\begin{aligned}&L(\mathbf{C},\mathbf{D}^{(1)},\cdots,\mathbf{D}^{(V)},\mathbf{E}^{(1)},\cdots,\mathbf{E}^{(V)},\mathbf{K})=\\ &\sum_{v=1}^{V}(\|\mathbf{E}^{(v)}\|_F^2+\lambda_D\|\mathbf{D}^{(v)}\|_2^2)+\lambda_C\|\mathbf{K}\|_*+\\ &\sum_{v=1}^{V}\langle\mathbf{Y}_1^{(v)},\mathbf{X}^{(v)}-\mathbf{X}^{(v)}(\mathbf{C}+\mathbf{D}^{(v)})-\mathbf{E}^{(v)}\rangle+\\ &\frac{\mu}{2}\sum_{v=1}^{V}\|\mathbf{X}^{(v)}-\mathbf{X}^{(v)}(\mathbf{C}+\mathbf{D}^{(v)})-\mathbf{E}^{(v)}\|_F^2+\\ &\langle\mathbf{Y}_2,\mathbf{C}-\mathbf{K}\rangle+\frac{\mu}{2}\|\mathbf{C}-\mathbf{K}\|_F^2,\end{aligned} \tag{3}$$

or equivalently,

$$\begin{aligned}&L(\mathbf{C},\mathbf{D}^{(1)},\cdots,\mathbf{D}^{(V)},\mathbf{E}^{(1)},\cdots,\mathbf{E}^{(V)},\mathbf{K})=\\ &\sum_{v=1}^{V}(\|\mathbf{E}^{(v)}\|_F^2+\lambda_D\|\mathbf{D}^{(v)}\|_2^2)+\lambda_C\|\mathbf{K}\|_*+\\ &\frac{\mu}{2}\sum_{v=1}^{V}\|\mathbf{X}^{(v)}(\mathbf{C}+\mathbf{D}^{(v)})+\mathbf{E}^{(v)}-(\mathbf{X}^{(v)}+\frac{\mathbf{Y}_1^{(v)}}{\mu})\|_F^2+\\ &\frac{\mu}{2}\|\mathbf{K}-(\mathbf{C}+\frac{\mathbf{Y}_2}{\mu})\|_F^2-\frac{1}{2\mu}(\sum_{v=1}^{V}\|\mathbf{Y}_1^{(v)}\|_F^2+\|\mathbf{Y}_2\|_F^2),\end{aligned} \tag{4}$$

where $\langle\cdot,\cdot\rangle$ is the standard Euclidean inner product of two matrices, $\{\mathbf{Y}_1^{(v)}\}_{v\in[V]},\mathbf{Y}_2$ are Lagrange multipliers and $\mu>0$ is a penalty parameter. To tackle this issue, we divide it into five sub-problems and alternatively optimize them by fixing the other variables.

Solving C: Obtain $\mathbf{C}$ by minimizing Eq. (4) while $\mathbf{K},\mathbf{D}^{(v)},\mathbf{E}^{(v)}(v\in[V])$ are fixed:

$$\begin{aligned}\min_{\mathbf{C}}&\|\mathbf{K}-(\mathbf{C}+\frac{\mathbf{Y}_2}{\mu})\|_F^2+\\ &\sum_{v=1}^{V}\|\mathbf{X}^{(v)}(\mathbf{C}+\mathbf{D}^{(v)})+\mathbf{E}^{(v)}-(\mathbf{X}^{(v)}+\frac{\mathbf{Y}_1^{(v)}}{\mu})\|_F^2.\end{aligned} \tag{5}$$

*Corresponding author.

Let the derivative of Eq. (5) be zero yields

$$\mathbf{C} = (\sum_{v=1}^{V}(\mathbf{X}^{(v)})^T\mathbf{X}^{(v)} + \mathbf{I})^{-1}\times$$
$$(\sum_{v=1}^{V}(\mathbf{X}^{(v)})^T(\mathbf{X}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu} - \mathbf{X}^{(v)}\mathbf{D}^{(v)} - \mathbf{E}^{(v)}) + \mathbf{K} - \frac{\mathbf{Y}_2}{\mu}), \tag{6}$$

where $\mathbf{I}$ denotes the identity matrix.

Solving $\mathbf{K}$: Obtain $\mathbf{K}$ by minimizing Eq. (4) while $\mathbf{C}, \mathbf{D}^{(v)}, \mathbf{E}^{(v)}(v \in [V])$ are fixed:

$$\min_{\mathbf{K}} \frac{\lambda_C}{\mu}\|\mathbf{K}\|_* + \frac{1}{2}\|\mathbf{K} - (\mathbf{C} + \frac{\mathbf{Y}_2}{\mu})\|_F^2. \tag{7}$$

Let $\mathbf{C} + \frac{\mathbf{Y}_2}{\mu} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the singular value decomposition, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Sigma} = diag\{\sigma_i\}$,$i = 1, 2, \cdots, r$ are singular values in descending order. Thanks to the singular value shrinkage operator (Cai, Candès, and Shen 2010), we get

$$\mathbf{K} = \mathcal{D}_{\frac{\lambda_C}{\mu}}(\mathbf{C} + \frac{\mathbf{Y}_2}{\mu}) = \mathbf{U}\mathcal{D}_{\frac{\lambda_C}{\mu}}(\mathbf{\Sigma})\mathbf{V}^T, \tag{8}$$

where $\mathcal{D}_\tau(\mathbf{\Sigma}) = diag\{(\sigma_i - \tau)_+\}$ and $t_+ = max(0, t)$.

Solving $\boldsymbol{D^{(v)}}, \boldsymbol{v} \in [\boldsymbol{V}]$: Obtain $\mathbf{D}^{(v)}$ by minimizing Eq. (4) while $\mathbf{C}, \mathbf{K}, \mathbf{E}^{(v)}(v \in [V])$ and $\mathbf{D}^{(w)}(w \neq v)$ are fixed:

$$\min_{\mathbf{D}^{(v)}} \lambda_D\|\mathbf{D}^{(v)}\|_2^2 + \frac{\mu}{2}\|\mathbf{X}^{(v)}(\mathbf{C} + \mathbf{D}^{(v)}) + \mathbf{E}^{(v)} - (\mathbf{X}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu})\|_F^2. \tag{9}$$

Let the derivative of Eq. (9) be zero gives

$$\mathbf{D}^{(v)} = (\mu(\mathbf{X}^{(v)})^T\mathbf{X}^{(v)} + 2\lambda_D\mathbf{I})^{-1}\times \mu(\mathbf{X}^{(v)})^T(\mathbf{X}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu} - \mathbf{X}^{(v)}\mathbf{C} - \mathbf{E}^{(v)}). \tag{10}$$

Solving $\boldsymbol{E^{(v)}}, \boldsymbol{v} \in [\boldsymbol{V}]$: Obtain $\mathbf{E}^{(v)}$ by minimizing Eq. (4) while $\mathbf{C}, \mathbf{K}, \mathbf{D}^{(v)}(v \in [V])$ and $\mathbf{E}^{(w)}(w \neq v)$ are fixed:

$$\min_{\mathbf{E}^{(v)}} \|\mathbf{E}^{(v)}\|_F^2 + \frac{\mu}{2}\|\mathbf{X}^{(v)}(\mathbf{C} + \mathbf{D}^{(v)}) + \mathbf{E}^{(v)} - (\mathbf{X}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu})\|_F^2. \tag{11}$$

Let the derivative of Eq. (11) be zero, we can get

$$\mathbf{E}^{(v)} = \frac{\mu}{2+\mu}(\mathbf{X}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu} - \mathbf{X}^{(v)}\mathbf{C} - \mathbf{X}^{(v)}\mathbf{D}^{(v)}). \tag{12}$$

Solving multipliers: Obtain $\mathbf{Y}_1^{(v)}(v \in [V]), \mathbf{Y}_2$ with every iteration t:

$$\mathbf{Y}_{1(t+1)}^{(v)} = \mathbf{Y}_{1(t)}^{(v)} + \mu(\mathbf{X}^{(v)} - \mathbf{X}^{(v)}(\mathbf{C} + \mathbf{D}^{(v)}) - \mathbf{E}^{(v)}),$$
$$\mathbf{Y}_{2(t+1)}^{(v)} = \mathbf{Y}_{2(t)}^{(v)} + \mu(\mathbf{C} - \mathbf{K}). \tag{13}$$

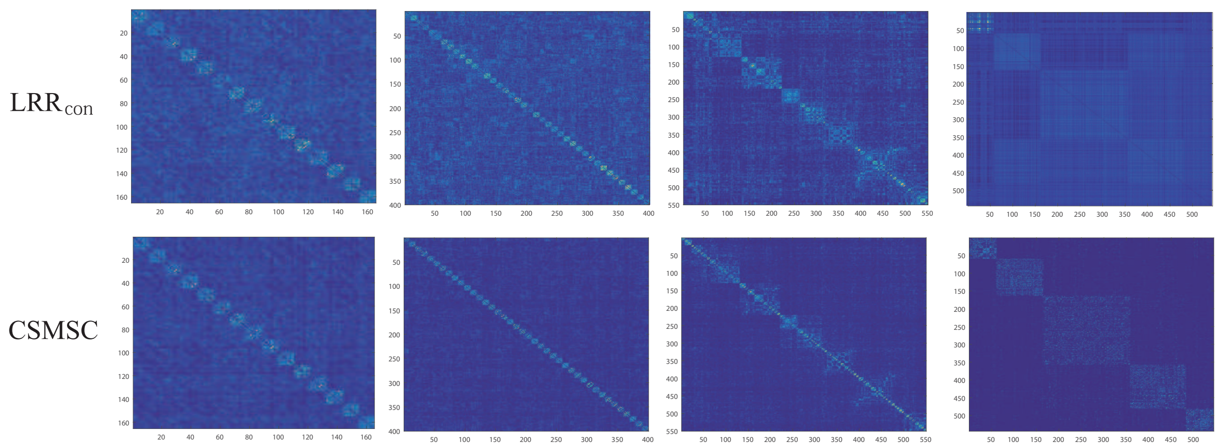
Experimental Results

Figure 1 further visualizes the affinity matrices produced by LRR_{con} and CSMSC on four datasets. As observed, the affinity matrices of our method plot the underlying clustering structures more clearly, which also validates the effectiveness of our method. In conclusion, our method better models the real-world data with common representation shared among views as well as modification of each view.

References

Cai, J.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20(4):1956–1982.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.



(a) Yale (b) ORL (c) Notting-Hill (d) BBCSport

Figure 1: Visualizing affinity matrices by $\mathrm{LRR}_{\mathrm{con}}$ (top) and CSMSC (bottom) on each dataset. As presented, our method produces much more clear structures (clusters) in the visualized affinity matrices. (BEST VIEWED IN COLOR)