Journal of Industrial and Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(1): 44–69.
DOI:10.3934/jimo.2026003
Received: 11 September 2025
Revised: 29 October 2025
Accepted: 06 November 2025
Published: 14 November 2025

Research article

Supplementary Material for "Group Sparsity-based Fusion Regularized Clustering: New Model and Convergent Algorithm"

Xiangru Xing[1], Lingchen Kong[1], Xin Wang[1,*] and Xianchao Xiu[2]

[1] School of Mathematics and Statistics, Beijing Jiaotong University, Beijing, 100044, China

[2] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China

* **Correspondence:** Email: xinwang2@bjtu.edu.cn.

I. Adaptive weights estimation

The choice of $\boldsymbol{\beta}$ significantly influences feature selection and clustering accuracy. Following the adaptive strategies proposed by [5], we present the adaptive updating procedure for $\tilde{\boldsymbol{\beta}}$ in Algorithm 2. Specifically, given a range of α and a fixed η, we first fit GSFRC with $\gamma = 1$ and uniform feature weights $\beta_j = 1$. We then identify the value of α that yields the desired number of clusters, obtaining the initial estimate $\tilde{\mathbf{X}}$.

We update the weights in two aspects. On one hand, adaptive feature weights are constructed based on the initial estimate $\tilde{\beta}_j = 1/(\|\tilde{\mathbf{X}}_{\cdot j}\|_2 + 0.01)$. These weights impose a strong penalty on noise features such that corresponding $\|\tilde{\mathbf{X}}_{\cdot j}\|_2$ tends to 0. On the other hand, to reduce the influence of noise features in $\boldsymbol{A}$ on similarity measurement, we apply min-max normalization to $\tilde{\boldsymbol{\beta}}$ to obtain $\boldsymbol{d}_\beta$ and use $\boldsymbol{d}_\beta$ as sample weights for each feature, i.e., $\|(\boldsymbol{A}_{i\cdot} - \boldsymbol{A}_{i'\cdot}) \odot \boldsymbol{d}_\beta\|_2^2$, which refines the computation of the original distance $\|\boldsymbol{A}_{i\cdot} - \boldsymbol{A}_{i'\cdot}\|_2^2$. Here, "$\odot$" denotes Hadamard product.

Algorithm 2 Adaptively adjusting $\boldsymbol{\beta}$ and ω_ι

Input: Data matrix $\mathbf{A}$, hyperparameter η
Parameters: Tuning parameter sequences for α and γ

Step 1. Fit GSFRC with $\gamma = 1, \beta_j = 1$ for $j = 1, 2, \ldots, d$, and a sequence of α.

Step 2. Select optimal α that yield the desired number of clusters; Obtain the preliminary estimate $\tilde{\mathbf{X}}$.

Step 3. Update the adaptive weights:

- Feature weights: $\tilde{\beta}_j = \dfrac{1}{\|\tilde{\mathbf{X}}_{\cdot j}\|_2 + 0.01}$ for $j = 1, 2, \ldots, d$.
- Fusion weights: $\tilde{\omega}_\iota = \begin{cases} \exp\Big(- \phi \Big\|(\boldsymbol{A}_{i\cdot} - \boldsymbol{A}_{i'\cdot}) \odot \boldsymbol{d}_\beta \Big\|_2^2 \Big), & \text{if } (i, i') \in \epsilon \\ 0, & \text{otherwise} \end{cases}$.

Step 4. Fit adaptive GSFRC with updated weights $\tilde{\boldsymbol{\beta}}$, $\tilde{\omega}_\iota$ and sequences of α and γ; Find optimal α and γ that achieve the desired number of clusters and features.

Output: Final cluster assignments, selected features, and refined estimate $\tilde{\mathbf{X}}$.

II. The proof of Lemma 4.2

Proof. 1) According to (4.2) and (4.8), we get

$$\begin{aligned} &\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t+1)}; \boldsymbol{\Upsilon}_2^{(t+1)}, \varepsilon^{(t+1)}) - \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t+1)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) \\ &= \langle \boldsymbol{P}^{(t+1)} - \boldsymbol{P}^{(t)}, \boldsymbol{D}\boldsymbol{X}^{(t+1)} - \boldsymbol{E}^{(t+1)} \rangle + \langle \boldsymbol{G}^{(t+1)} - \boldsymbol{G}^{(t)}, \boldsymbol{X}^{(t+1)} - \boldsymbol{F}^{(t+1)} \rangle \\ &+ \langle \bar{\boldsymbol{G}}^{(t+1)} - \bar{\boldsymbol{G}}^{(t)}, \boldsymbol{X}^{(t+1)} - \bar{\boldsymbol{F}}^{(t+1)} \rangle + \varepsilon^{(t+1)} - \varepsilon^{(t)} \\ &= \frac{1}{\tau}\|\boldsymbol{\Upsilon}_2^{(t+1)} - \boldsymbol{\Upsilon}_2^{(t)}\|_F^2 + \varepsilon^{(t+1)} - \varepsilon^{(t)} \\ &\leq \frac{\theta}{\tau}\|\boldsymbol{X}^{(t+1)} - \boldsymbol{X}^{(t)}\|_F^2, \end{aligned} \tag{6.1}$$

where the second "=" and the first "≤" hold via (4.7) and Assumption 4.1, respectively.

Let

$$\begin{aligned} Q_t(\boldsymbol{X}) = &\frac{1}{2}\|\boldsymbol{X} - \boldsymbol{A}\|_F^2 + \frac{\tau}{2}\|\boldsymbol{D}\boldsymbol{X} - \boldsymbol{E}^{(t+1)} + \frac{1}{\tau}\boldsymbol{P}^{(t)}\|_F^2 \\ &+ \frac{\tau}{2}\|\boldsymbol{X} - \boldsymbol{F}^{(t+1)} + \frac{1}{\tau}\boldsymbol{G}^{(t)}\|_F^2 + \frac{\tau}{2}\|\boldsymbol{X} - \bar{\boldsymbol{F}}^{(t+1)} + \frac{1}{\tau}\bar{\boldsymbol{G}}^{(t)}\|_F^2. \end{aligned}$$

Since $\boldsymbol{X}^{(t+1)}$ is the optimal solution of $\boldsymbol{X}$-subproblem, $\nabla Q_t(\boldsymbol{X}^{(t+1)}) = 0$. Combined with the strong convexity of $Q_t(\boldsymbol{X})$, we derive that

$$\begin{aligned} &\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t+1)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) - \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) \\ &= Q_t(\boldsymbol{X}^{(t+1)}) - Q_t(\boldsymbol{X}^{(t)}) \\ &\leq -\frac{2\tau + 1}{2}\|\boldsymbol{X}^{(t+1)} - \boldsymbol{X}^{(t)}\|_F^2. \end{aligned} \tag{6.2}$$

Similarly, by the optimality of $\boldsymbol{E}^{(t+1)}$ for the $\boldsymbol{E}$-subproblem, $\boldsymbol{F}^{(t+1)}$ for the $\boldsymbol{F}$-subproblem, and $\bar{\boldsymbol{F}}^{(t+1)}$ for the $\bar{\boldsymbol{F}}$-subproblem, we get

$$\begin{aligned} & g(\boldsymbol{E}^{(t+1)}) + \frac{\tau}{2}\|\boldsymbol{D}\boldsymbol{X}^{(t)} - \boldsymbol{E}^{(t+1)} + \frac{1}{\tau}\boldsymbol{P}^{(t)}\|_F^2 \leq g(\boldsymbol{E}^{(t)}) + \frac{\tau}{2}\|\boldsymbol{D}\boldsymbol{X}^{(t)} - \boldsymbol{E}^{(t)} + \frac{1}{\tau}\boldsymbol{P}^{(t)}\|_F^2, \\ & r_1(\boldsymbol{F}^{(t+1)}) + \frac{\tau}{2}\|\boldsymbol{X}^{(t)} - \boldsymbol{F}^{(t+1)} + \frac{1}{\tau}\boldsymbol{G}^{(t)}\|_F^2 \leq r_1(\boldsymbol{F}^{(t)}) + \frac{\tau}{2}\|\boldsymbol{X}^{(t)} - \boldsymbol{F}^{(t)} + \frac{1}{\tau}\boldsymbol{G}^{(t)}\|_F^2, \\ & r_2(\bar{\boldsymbol{F}}^{(t+1)}) + \frac{\tau}{2}\|\boldsymbol{X}^{(t)} - \bar{\boldsymbol{F}}^{(t+1)} + \frac{1}{\tau}\bar{\boldsymbol{G}}^{(t)}\|_F^2 \leq r_2(\bar{\boldsymbol{F}}^{(t)}) + \frac{\tau}{2}\|\boldsymbol{X}^{(t)} - \bar{\boldsymbol{F}}^{(t)} + \frac{1}{\tau}\bar{\boldsymbol{G}}^{(t)}\|_F^2. \end{aligned} \tag{6.3}$$

Using (6.3), we have

$$\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) - \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) \leq 0. \tag{6.4}$$

Summing both sides of (6.1), (6.2), and (6.4), we can obtain

$$\begin{aligned} & \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t+1)}; \boldsymbol{\Upsilon}_2^{(t+1)}, \varepsilon^{(t+1)}) - \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) \\ & \leq -\bar{w}\|\boldsymbol{X}^{(t+1)} - \boldsymbol{X}^{(t)}\|_F^2 \\ & < 0, \end{aligned}$$

where $\bar{w} = \frac{2\tau+1}{2} - \frac{\theta}{\tau}$.

2) Based on the below optimality conditions of the $\boldsymbol{E}$-subproblem, $\boldsymbol{F}$-subproblem, $\bar{\boldsymbol{F}}$-subproblem, and $\boldsymbol{X}$-subproblem

$$\begin{cases} 0 \in \partial g(\boldsymbol{E}^{(t)}) - \tau(\boldsymbol{D}\boldsymbol{X}^{(t-1)} - \boldsymbol{E}^{(t)} + \frac{1}{\tau}\boldsymbol{P}^{(t-1)}), \\ 0 \in \partial r_1(\boldsymbol{F}^{(t)}) - \tau(\boldsymbol{X}^{(t-1)} - \boldsymbol{F}^{(t)} + \frac{1}{\tau}\boldsymbol{G}^{(t-1)}), \\ 0 \in \partial r_2(\bar{\boldsymbol{F}}^{(t)}) - \tau(\boldsymbol{X}^{(t-1)} - \bar{\boldsymbol{F}}^{(t)} + \frac{1}{\tau}\bar{\boldsymbol{G}}^{(t-1)}), \\ 0 = \boldsymbol{X}^{(t)} - \boldsymbol{A} + \boldsymbol{D}^\top\boldsymbol{P}^{(t)} + \boldsymbol{G}^{(t)} + \bar{\boldsymbol{G}}^{(t)}, \end{cases} \tag{6.5}$$

and the subdifferentials of $\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)})$ with respect to $\boldsymbol{E}^{(t)}$, $\boldsymbol{F}^{(t)}$, $\bar{\boldsymbol{E}}^{(t)}$, $\boldsymbol{X}^{(t)}$, $\boldsymbol{P}^{(t)}$, $\boldsymbol{G}^{(t)}$, and $\bar{\boldsymbol{G}}^{(t)}$

$$\begin{cases} \partial_{\boldsymbol{E}}\hat{\mathcal{L}}_\tau &= \partial g(\boldsymbol{E}^{(t)}) - 2\boldsymbol{P}^{(t)} + \boldsymbol{P}^{(t-1)}, \\ \partial_{\boldsymbol{F}}\hat{\mathcal{L}}_\tau &= \partial r_1(\boldsymbol{F}^{(t)}) - 2\boldsymbol{G}^{(t)} + \boldsymbol{G}^{(t-1)}, \\ \partial_{\bar{\boldsymbol{F}}}\hat{\mathcal{L}}_\tau &= \partial r_2(\bar{\boldsymbol{F}}^{(t)}) - 2\bar{\boldsymbol{G}}^{(t)} + \bar{\boldsymbol{G}}^{(t-1)}, \\ \partial_{\boldsymbol{X}}\hat{\mathcal{L}}_\tau &= \boldsymbol{X}^{(t)} - \boldsymbol{A} + \boldsymbol{D}^\top(2\boldsymbol{P}^{(t)} - \boldsymbol{P}^{(t-1)}) \\ & \quad +2\boldsymbol{G}^{(t)} - \boldsymbol{G}^{(t-1)} + 2\bar{\boldsymbol{G}}^{(t)} - \bar{\boldsymbol{G}}^{(t-1)}, \\ \partial_{\boldsymbol{\Upsilon}_2}\hat{\mathcal{L}}_\tau &= \frac{1}{\tau}(\boldsymbol{\Upsilon}_2^{(t)} - \boldsymbol{\Upsilon}_2^{(t-1)}), \end{cases} \tag{6.6}$$

we have

$$\begin{cases} \partial_{\boldsymbol{E}}\hat{\mathcal{L}}_\tau \ni \tau\boldsymbol{D}(\boldsymbol{X}^{(t-1)} - \boldsymbol{X}^{(t)}) + \boldsymbol{P}^{(t-1)} - \boldsymbol{P}^{(t)}, \\ \partial_{\boldsymbol{F}}\hat{\mathcal{L}}_\tau \ni \tau(\boldsymbol{X}^{(t-1)} - \boldsymbol{X}^{(t)}) + \boldsymbol{G}^{(t-1)} - \boldsymbol{G}^{(t)}, \\ \partial_{\bar{\boldsymbol{F}}}\hat{\mathcal{L}}_\tau \ni \tau(\boldsymbol{X}^{(t-1)} - \boldsymbol{X}^{(t)}) + \bar{\boldsymbol{G}}^{(t-1)} - \bar{\boldsymbol{G}}^{(t)}, \\ \partial_{\boldsymbol{X}}\hat{\mathcal{L}}_\tau = \boldsymbol{D}^\top(\boldsymbol{P}^{(t)} - \boldsymbol{P}^{(t-1)}) + \boldsymbol{G}^{(t)} - \boldsymbol{G}^{(t-1)} + \bar{\boldsymbol{G}}^{(t)} - \bar{\boldsymbol{G}}^{(t-1)}, \end{cases} \tag{6.7}$$

and

$$\boldsymbol{A} - \boldsymbol{X}^{(t)} = \boldsymbol{D}^\top \boldsymbol{P}^{(t)} + \boldsymbol{G}^{(t)} + \bar{\boldsymbol{G}}^{(t)}. \tag{6.8}$$

Thus,

$$\begin{aligned}
&\mathrm{dist}(0, \partial \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)})) \\
&\leq \|(\overrightarrow{\partial_{\boldsymbol{E}} \hat{\mathcal{L}}_\tau}, \overrightarrow{\partial_{\boldsymbol{F}} \hat{\mathcal{L}}_\tau}, \overrightarrow{\partial_{\bar{\boldsymbol{F}}} \hat{\mathcal{L}}_\tau}, \overrightarrow{\partial_{\boldsymbol{X}} \hat{\mathcal{L}}_\tau}, \overrightarrow{\partial_{\boldsymbol{\Upsilon}_2} \hat{\mathcal{L}}_\tau})\|_F \\
&\leq (2\tau + \tau\sqrt{\lambda_{max}(\boldsymbol{D}^\top \boldsymbol{D})} + 1)\|\boldsymbol{X}^{(t)} - \boldsymbol{X}^{(t-1)}\|_F + (\frac{1}{\tau} + 1)\|\boldsymbol{\Upsilon}_2^{(t)} - \boldsymbol{\Upsilon}_2^{(t-1)}\|_F \\
&\leq (2\tau + \tau\sqrt{\lambda_{max}(\boldsymbol{D}^\top \boldsymbol{D})} + 1 + \frac{\sqrt{\theta}}{\tau} + \sqrt{\theta})\|\boldsymbol{X}^{(t)} - \boldsymbol{X}^{(t-1)}\|_F \\
&+ (\frac{\sqrt{\tau}}{\tau} + \sqrt{\tau})\sqrt{(\varepsilon^{(t-1)} - \varepsilon^{(t)})} \\
&\leq \tilde{w}(\|\boldsymbol{X}^{(t)} - \boldsymbol{X}^{(t-1)}\|_F + \sqrt{\varepsilon^{(t-1)} - \varepsilon^{(t)}}),
\end{aligned}$$

where the second "≤" holds via (6.6)-(6.8) and $\sum_{i=1}^n (a_i)^2 \leq (\sum_{i=1}^n a_i)^2$ when $a_i \geq 0$, and the third "≤" holds via Assumption 4.1 and $\sum_{i=1}^n (a_i)^2 \leq (\sum_{i=1}^n a_i)^2$ when $a_i \geq 0$. □

III. The proof of Lemma 4.3

Proof. 1) From Lemma 4.2, we get

$$\begin{aligned}
&\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(1)}, \boldsymbol{X}^{(1)}; \boldsymbol{\Upsilon}_2^{(1)}, \varepsilon^{(1)}) \geq \cdots \geq \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) \\
&= \frac{1}{2}\|\boldsymbol{X}^{(t)} - \boldsymbol{A}\|_F^2 + g(\boldsymbol{E}^{(t)}) + r_1(\boldsymbol{F}^{(t)}) + r_2(\bar{\boldsymbol{F}}^{(t)}) \\
&+ \frac{\tau}{2}\|\boldsymbol{D}\boldsymbol{X}^{(t)} - \boldsymbol{E}^{(t)} + \frac{1}{\tau}\boldsymbol{P}^{(t)}\|_F^2 + \frac{\tau}{2}\|\boldsymbol{X}^{(t)} - \boldsymbol{F}^{(t)} + \frac{1}{\tau}\boldsymbol{G}^{(t)}\|_F^2 \\
&+ \frac{\tau}{2}\|\boldsymbol{X}^{(t)} - \bar{\boldsymbol{F}}^{(t)} + \frac{1}{\tau}\bar{\boldsymbol{G}}^{(t)}\|_F^2 - \frac{1}{2\tau}\|\boldsymbol{\Upsilon}_2^{(t)}\|_F^2 + \varepsilon^{(t)}.
\end{aligned}$$

Then, $\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \boldsymbol{\Upsilon}_2^{(t)}, \varepsilon^{(t)})$ is coercive, bounded from below, and lower semi-continuous since the objective function in (1.4) is coercive, bounded from below, and lower semi-continuous, $\hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(1)}, \boldsymbol{X}^{(1)}; \boldsymbol{\Upsilon}_2^{(1)}, \varepsilon^{(1)}) < +\infty$, and Assumption 4.1 holds. Further, it is obvious that the generated sequences $\{\boldsymbol{\Upsilon}_1^{(t)}\}_{t=1}^{+\infty}$ and $\{\boldsymbol{X}^{(t)}\}_{t=1}^{+\infty}$ are bounded.

2) In accordance with the boundedness of $\{(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{\Upsilon}_2^{(t)})\}_{t=1}^{+\infty}$ and the Bolzano-Weierstrass theorem, there exists at least one cluster point $\tilde{\boldsymbol{\Upsilon}} = (\tilde{\boldsymbol{\Upsilon}}_1, \tilde{\boldsymbol{X}}; \tilde{\boldsymbol{\Upsilon}}_2)$ and a subsequence $\{(\boldsymbol{\Upsilon}_1^{(t_j)}, \boldsymbol{X}^{(t_j)}, \boldsymbol{\Upsilon}_2^{(t_j)})\}_{t=1}^{+\infty} \subseteq \{(\boldsymbol{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}, \boldsymbol{\Upsilon}_2^{(t)})\}_{t=1}^{+\infty}$ such that

$$\lim_{j \to +\infty} (\boldsymbol{\Upsilon}_1^{(t_j)}, \boldsymbol{X}^{(t_j)}; \boldsymbol{\Upsilon}_2^{(t_j)}) = \tilde{\boldsymbol{\Upsilon}}, \tag{6.9}$$

where $\tilde{\boldsymbol{\Upsilon}}_1 = (\tilde{\boldsymbol{E}}, \tilde{\boldsymbol{F}}, \tilde{\bar{\boldsymbol{F}}})$ and $\tilde{\boldsymbol{\Upsilon}}_2 = (\tilde{\boldsymbol{P}}, \tilde{\boldsymbol{G}}, \tilde{\bar{\boldsymbol{G}}})$.

From the properness and the lower semi-continuity of $\hat{\mathcal{L}}_\tau(\cdot)$, we can get

$$\liminf_{j \to +\infty} \hat{\mathcal{L}}_\tau(\boldsymbol{\Upsilon}_1^{(t_j)}, \boldsymbol{X}^{(t_j)}; \boldsymbol{\Upsilon}_2^{(t_j)}, \varepsilon^{(t_j)}) \geq \mathcal{L}_\tau(\tilde{\boldsymbol{\Upsilon}}) > -\infty.$$

Using Lemma 4.2, we have

$$\begin{aligned}
&\sum_{t=1}^{t_{j-1}} \|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 \\
&\leq \frac{1}{\bar{w}} \sum_{t=1}^{t_{j-1}} (\hat{\mathcal{L}}_\tau(\mathbf{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \mathbf{\Upsilon}_2^{(t)}, \varepsilon^{(t)}) - \hat{\mathcal{L}}_\tau(\mathbf{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t+1)}; \mathbf{\Upsilon}_2^{(t+1)}, \varepsilon^{(t+1)})) \\
&\leq \frac{1}{\bar{w}} (\hat{\mathcal{L}}_\tau(\mathbf{\Upsilon}_1^{(1)}, \boldsymbol{X}^{(1)}; \mathbf{\Upsilon}_2^{(1)}, \varepsilon^{(1)}) - \hat{\mathcal{L}}_\tau(\mathbf{\Upsilon}_1^{(t_j)}, \boldsymbol{X}^{(t_j)}; \mathbf{\Upsilon}_2^{(t_j)}, \varepsilon^{(t_j)})),
\end{aligned}$$

which indicates that

$$\sum_{t=1}^{+\infty} \|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 < +\infty \text{ as } j \to +\infty. \tag{6.10}$$

Subsequently, it holds from Assumption 4.1 that

$$\begin{aligned}
\sum_{t=1}^{+\infty} \|\mathbf{\Upsilon}_2^{(t+1)} - \mathbf{\Upsilon}_2^{(t)}\|_F^2 &= \sum_{t=1}^{+\infty} \theta \|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 + \sum_{t=1}^{+\infty} \tau(\varepsilon^{(t+1)} - \varepsilon^{(t)}) \\
&= \sum_{t=1}^{+\infty} \theta \|\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}\|_F^2 - \tau\varepsilon^{(1)} \\
&< +\infty.
\end{aligned} \tag{6.11}$$

Moreover, we obtain with (4.7) that

$$\sum_{t=1}^{+\infty} \|\mathbf{\Upsilon}_1^{(t+1)} - \mathbf{\Upsilon}_1^{(t)}\|_F^2 \leq \frac{1}{\tau} \sum_{t=1}^{+\infty} \|\mathbf{\Upsilon}_2^{(t+1)} - \mathbf{\Upsilon}_2^{(t)}\|_F^2 < +\infty. \tag{6.12}$$

Therefore, it follows from (6.9)-(6.12), that

$$\lim_{t \to +\infty} \left\| (\mathbf{\Upsilon}_1^{(t+1)}, \boldsymbol{X}^{(t+1)}; \mathbf{\Upsilon}_2^{(t+1)}) - (\mathbf{\Upsilon}_1^{(t)}, \boldsymbol{X}^{(t)}; \mathbf{\Upsilon}_2^{(t)}) \right\|_F = 0.$$

□

IV. Optimal parameter settings

We compare the accuracy of different methods on raw, unprocessed real-world datasets in Section 5. The optimal parameter values for each method are listed in Table 6, with "-" indicating the absence of a corresponding parameter. It is important to note that while the optimal parameters may lie within a range, we have reported only a single representative value for conciseness.

Table 6. Hyperparameter settings of different clustering methods on real-world datasets.

Datasets	Methods	α	γ	η	ϕ	k	τ
Authors	K-means	-	-	-	-	-	-
	CC	8	-	-	10^{-5}	3	-
	SCC	3756	1	-	5	3	1
	SGLCC	5756	0.7	2^{-8}	5	3	0.4
	ERC	150	-	-	10^{-5}	3	-
	GSFRC	5255	0.01	2^{-8}	5	3	0.4
Lung-discrete	K-means	-	-	-	-	-	-
	CC	110	-	-	10^{-3}	3	-
	SCC	7809	1	-	5	3	1
	SGLCC	5961	1	2^{-8}	5	3	0.5
	ERC	11.028	-	-	10^{-5}	4	-
	GSFRC	5250	6	2^{-6}	5	3	0.5
GLIOMA	K-means	-	-	-	-	-	-
	CC	19	-	-	10^{-5}	3	-
	SCC	937	0.7	-	5	6	1
	SGLCC	909	0.03	2^{-7}	5	6	0.3
	ERC	6	-	-	10^{-5}	4	-
	GSFRC	942	0.05	2^{-7}	5	6	0.3
Brain	K-means	-	-	-	-	-	-
	CC	48	-	-	10^{-5}	3	-
	SCC	6300	0.04	-	5	3	1
	SGLCC	5488	0.01	2^{-4}	5	3	0.8
	ERC	25	-	-	10^{-5}	3	-
	GSFRC	5224	0.004	2^{-4}	5	3	0.8