Plug-and-Play Incomplete Multi-View Clustering via Janus-Faced Affinity Learning with Topology Harmonization

Supplementary Material

1. Supporting Derivations

This section provides a detailed introduction to the derivation process for solving the designed objective function (5).

Given the non-convex nature of the problem, we address it by iteratively optimizing for subsets of variables.

1.1. $\mathbf{C}_v$ Optimization

With fixed $\mathbf{Q}_v$, $\mathbf{H}$, $\mathbf{H}_v$, θ_v and β_v, the objective function (5) is equivalently expressed as

$$\min_{\mathbf{C}_v} \sum_{v=1}^{V} \beta_v^2 \left\| \mathbf{D}_v\mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v \right\|_F^2 \quad \text{s.t. } \mathbf{C}_v^\top \mathbf{C}_v = \mathbf{I}_m. \tag{29}$$

Then, by virtue of the independence between views, we can obtain

$$\min_{\mathbf{C}_v} \left\| \mathbf{D}_v\mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v \right\|_F^2 \quad \text{s.t. } \mathbf{C}_v^\top \mathbf{C}_v = \mathbf{I}_m. \tag{30}$$

Through trace expanding and orthogonality of $\mathbf{C}_v$, we can derive

$$\begin{aligned}
&\min_{\mathbf{C}_v} \left\| \mathbf{D}_v\mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v \right\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{C}_v} \left\| \mathbf{D}_v\mathbf{E}_v - \mathbf{C}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right) \right\|_F^2 \Leftrightarrow \\
&\min_{\mathbf{C}_v} \mathrm{Tr}\Big(\left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right) \cdot \\
&\qquad \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \mathbf{C}_v \\
&\qquad - 2\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \Big) \Leftrightarrow \\
&\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \right).
\end{aligned} \tag{31}$$

Assuming that the result of the singular value decomposition of $\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top$ is $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, in conjunction with the trace cycling property, we have the following equation,

$$\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \right) = \max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{V}^\top \mathbf{C}_v^\top \mathbf{U}\mathbf{\Sigma} \right). \tag{32}$$

Note that the diagonal elements of $\mathbf{\Sigma}$ are non-negative and that $\mathbf{V}^\top \mathbf{C}_v^\top \mathbf{U}$ is orthogonal. These properties ensure

$$\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \right) \leq \max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{\Sigma} \right), \tag{33}$$

where the equality holds when $\mathbf{V}^\top \mathbf{C}_v^\top \mathbf{U}$ equals to identity matrix. Therefore, the optimal $\mathbf{C}_v$ is $\mathbf{U}\mathbf{V}^\top$.

During optimizing $\mathbf{C}_v$, it requires constructing the item $\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top$. Combined with the matrix size, we have that this will need $\mathcal{O}(d_v n n_v + m^2 n + m n n_v + d_v n_v m)$ computing overhead. When the incomplete percentage is low, i.e., n_v is close to n, this complexity will approach $\mathcal{O}(n^2)$. As a result, this impedes large-scale deployment. To reduce it, we observe that

$$\begin{aligned}
&\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \right) \\
&\Leftrightarrow \\
&\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top \left(\theta_v \mathbf{Q}_v \mathbf{H} - (\theta_v - 1)\mathbf{H}_v \right)^\top \mathbf{C}_v^\top \right).
\end{aligned} \tag{34}$$

Further, in conjunction with the characteristic that each column of $\mathbf{E}_v$ contains a single 1 while other elements are all 0, we have that $\mathbf{E}_v\mathbf{E}_v^\top$ is a diagonal matrix. On the basis of this, we can derive that $\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top$ can be equivalently expressed as $\mathbf{D}_v \odot \mathbf{B}_v$, in which $\mathbf{B}_v$ is equal to

$$\mathbf{1}_{d_v} \cdot \left[\sum_{j=1}^{n_v} [\mathbf{E}_v]_{1,j}; \cdots; \sum_{j=1}^{n_v} [\mathbf{E}_v]_{n,j} \right]_n^\top.$$

Therefore, we have

$$\begin{aligned}
&\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v\mathbf{E}_v \left(\theta_v \mathbf{Q}_v \mathbf{H}\mathbf{E}_v - (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v \right)^\top \mathbf{C}_v^\top \right) \\
&\Leftrightarrow \\
&\max_{\mathbf{C}_v} \mathrm{Tr}\left(\mathbf{D}_v \odot \mathbf{B}_v \left(\theta_v \mathbf{Q}_v \mathbf{H} - (\theta_v - 1)\mathbf{H}_v \right)^\top \mathbf{C}_v^\top \right).
\end{aligned} \tag{35}$$

Constructing $\mathbf{D}_v \odot \mathbf{B}_v \left(\theta_v \mathbf{Q}_v \mathbf{H} - (\theta_v - 1)\mathbf{H}_v \right)^\top$ will need $\mathcal{O}(m^2 n + d_v n m)$ overhead. Therefore, we can conduct singular value decomposition on $\mathbf{D}_v \odot \mathbf{B}_v \left(\theta_v \mathbf{Q}_v \mathbf{H} - (\theta_v - 1)\mathbf{H}_v \right)^\top$ to generate the singular matrices $\mathbf{U}$ and $\mathbf{V}$, and then build the solution of $\mathbf{C}_v$ via $\mathbf{U}$ multiplying $\mathbf{V}^\top$. The total computing overhead about optimizing $\mathbf{C}_v$ is $\mathcal{O}(m^2 n + d_v n m + d_v m^2)$, which is linear to n since d_v is a constant and m is generally smaller than n.

1.2. $\mathbf{Q}_v$ Optimization

With fixed $\mathbf{C}_v$, $\mathbf{H}$, $\mathbf{H}_v$, θ_v and β_v, the objective function (5) becomes

$$\min_{\mathbf{Q}_v} \sum_{v=1}^{V} \beta_v^2 \left\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right\|_F^2 \quad \text{s.t. } \mathbf{Q}_v^\top \mathbf{1}_m = \mathbf{1}_m, \mathbf{Q}_v\mathbf{1}_m = \mathbf{1}_m, \mathbf{Q}_v \in \{0,1\}. \tag{36}$$

In conjunction with the view independence, we can further have

$$\min_{\mathbf{Q}_v} \left\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right\|_F^2 \quad \text{s.t. } \mathbf{Q}_v^\top \mathbf{1}_m = \mathbf{1}_m, \mathbf{Q}_v\mathbf{1}_m = \mathbf{1}_m, \mathbf{Q}_v \in \{0,1\}. \tag{37}$$

Combining norm expanding and trace looping, we can obtain

$$\begin{aligned}
&\min_{\mathbf{Q}_v} \left\|\mathbf{D}_v\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right\|_F^2 \\
&\Leftrightarrow \\
&\min_{\mathbf{Q}_v} \operatorname{Tr}\Big(\theta_v^2\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}^\top\mathbf{Q}_v^\top\mathbf{Q}_v \\
&\qquad - 2\theta_v\mathbf{H}\mathbf{E}_v\left(\mathbf{D}_v\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right)^\top\mathbf{C}_v\mathbf{Q}_v\Big).
\end{aligned} \tag{38}$$

Subsequently, according to the characteristics of feasible domain, we can obtain $\mathbf{Q}_v^\top\mathbf{Q}_v = \mathbf{I}$. Therefore, we have

$$\begin{aligned}
&\min_{\mathbf{Q}_v} \left\|\mathbf{D}_v\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right\|_F^2 \\
&\Leftrightarrow \\
&\max_{\mathbf{Q}_v} \operatorname{Tr}\left(\mathbf{H}\mathbf{E}_v\left(\mathbf{D}_v\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right)^\top\mathbf{C}_v\mathbf{Q}_v\right).
\end{aligned} \tag{39}$$

By the matrix-vectorization function Vec, we can derive

$$\begin{aligned}
&\max_{\mathbf{Q}_v} \operatorname{Tr}\left(\mathbf{H}\mathbf{E}_v\left(\mathbf{D}_v\mathbf{E}_v + \widehat{\mathbf{A}}_v\right)^\top\mathbf{C}_v\mathbf{Q}_v\right) \\
&\Leftrightarrow \\
&\max_{\mathbf{Q}_v} \left(\operatorname{Vec}\left(\left(\mathbf{H}\mathbf{E}_v\left(\widehat{\mathbf{G}}_v + \widehat{\mathbf{A}}_v\right)^\top\mathbf{C}_v\right)^\top\right)\right)^\top \operatorname{Vec}\left(\mathbf{Q}_v\right),
\end{aligned} \tag{40}$$

where $\widehat{\mathbf{G}}_v = \mathbf{D}_v\mathbf{E}_v$ and $\widehat{\mathbf{A}}_v = (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v$.

Then, we need to build the item $\mathbf{H}\mathbf{E}_v\left(\mathbf{D}_v\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right)^\top\mathbf{C}_v$, which requires $\mathcal{O}(d_vnn_v + d_vmn + mnn_v + mn_vd_v + m^2d_v)$ computing overhead. It approaches $\mathcal{O}(n^2)$. To reduce it, we observe that

$$\begin{aligned}
&\max_{\mathbf{Q}_v} \left(\operatorname{Vec}\left(\left(\mathbf{H}\mathbf{E}_v\left(\widehat{\mathbf{G}}_v + \widehat{\mathbf{A}}_v\right)^\top\mathbf{C}_v\right)^\top\right)\right)^\top \operatorname{Vec}\left(\mathbf{Q}_v\right) \\
&\Leftrightarrow \\
&\max_{\mathbf{Q}_v} \left(\operatorname{Vec}\left(\mathbf{C}_v^\top\left(\mathbf{D}_v\mathbf{E}_v + \widehat{\mathbf{A}}_v\right)\mathbf{E}_v^\top\mathbf{H}^\top\right)\right)^\top \operatorname{Vec}\left(\mathbf{Q}_v\right) \\
&\Leftrightarrow \\
&\max_{\mathbf{Q}_v} \left(\operatorname{Vec}\left(\left(\mathbf{C}_v^\top\mathbf{D}_v\mathbf{E}_v + \widehat{\mathbf{B}}_v\right)\mathbf{E}_v^\top\mathbf{H}^\top\right)\right)^\top \operatorname{Vec}\left(\mathbf{Q}_v\right) \\
&\Leftrightarrow \\
&\max_{\mathbf{Q}_v} \left(\operatorname{Vec}\left(\left(\mathbf{C}_v^\top\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top + \widehat{\mathbf{B}}_v\mathbf{E}_v^\top\right)\mathbf{H}^\top\right)\right)^\top \operatorname{Vec}\left(\mathbf{Q}_v\right),
\end{aligned} \tag{41}$$

where $\widehat{\mathbf{B}}_v = (\theta_v - 1)\mathbf{H}_v\mathbf{E}_v$.
Following the strategy handling the item $\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top$ in optimizing $\mathbf{C}_v$, we can equivalently express $\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top$ as $\mathbf{H}_v \odot \mathbf{G}_v$. $\mathbf{G}_v$ is composed of

$$\mathbf{1}_m \left[\sum_{j=1}^{n_v} [\mathbf{E}_v]_{1,j}; \cdots; \sum_{j=1}^{n_v} [\mathbf{E}_v]_{n,j}\right]_n^\top.$$

Consequently, we can get

$$\begin{aligned}
&\max_{\mathbf{Q}_v} \operatorname{Tr}\left(\mathbf{H}\mathbf{E}_v\left(\mathbf{D}_v\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right)^\top\mathbf{C}_v\mathbf{Q}_v\right) \\
&\Leftrightarrow \\
&\max_{\mathbf{Q}_v} \left(\operatorname{Vec}\left(\left(\mathbf{C}_v^\top\left(\mathbf{D}_v \odot \mathbf{B}_v\right) + \widehat{\mathbf{C}}_v\right)\mathbf{H}^\top\right)\right)^\top \operatorname{Vec}\left(\mathbf{Q}_v\right).
\end{aligned} \tag{42}$$

where $\widehat{\mathbf{C}}_v = (\theta_v - 1)\left(\mathbf{H}_v \odot \mathbf{G}_v\right)$.
Constructing $\left(\mathbf{C}_v^\top\left(\mathbf{D}_v \odot \mathbf{B}_v\right) + (\theta_v - 1)\left(\mathbf{H}_v \odot \mathbf{G}_v\right)\right)\mathbf{H}^\top$ only needs $\mathcal{O}(md_vn + m^2n)$ computing overhead. Additionally, according to the feasible domain, we have that this optimization problem is a binary programming, and can be solved efficiently by existing software packages. Therefore, the computational complexity of optimizing the variable $\mathbf{Q}_v$ is $\mathcal{O}(n)$.

1.3. $\mathbf{H}$ Optimization

With fixed $\mathbf{C}_v$, $\mathbf{Q}_v$, $\mathbf{H}_v$, θ_v and β_v, the objective function (5) simplifies to

$$\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \left\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right\|_F^2 \quad \text{s.t. } \mathbf{H}^\top\mathbf{1}_m = \mathbf{1}_n, \mathbf{H} \geq 0. \tag{43}$$

Unfolding by trace and removing irrelevant items yield

$$
\begin{aligned}
&\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \\
&\Leftrightarrow \\
&\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \left\| \mathbf{D}_v \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v \right\|_F^2 \\
&\Leftrightarrow \\
&\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \operatorname{Tr} \Big(\theta_v^2 \mathbf{E}_v^\top \mathbf{H}^\top \mathbf{Q}_v^\top \mathbf{C}_v^\top \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v \\
&\qquad - 2\theta_v \left(\mathbf{D}_v \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right) \mathbf{E}_v^\top \mathbf{H}^\top \mathbf{Q}_v^\top \mathbf{C}_v^\top \Big) \\
&\Leftrightarrow \\
&\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \operatorname{Tr} \Big(\mathbf{H}^\top \theta_v^2 \mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top \\
&\qquad - 2\theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top + (\theta_v - 1) \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top \right) \mathbf{H}^\top \Big).
\end{aligned} \tag{44}
$$

Combined with the fact that constraints $\mathbf{H}^\top \mathbf{1}_m = \mathbf{1}_n$ and $\mathbf{H} \geq 0$ aim at limiting the columns of $\mathbf{H}$, we can update $\mathbf{H}$ by column. As a result, we have

$$
\begin{aligned}
&\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \operatorname{Tr} \Big(\mathbf{H}^\top \theta_v^2 \mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top \\
&\qquad - 2\theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top + (\theta_v - 1) \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top \right) \mathbf{H}^\top \Big) \\
&\Leftrightarrow \\
&\min_{\mathbf{H}_{:,j}} \sum_{v=1}^{V} \beta_v^2 \Big(\mathbf{H}_{:,j}^\top \theta_v^2 \mathbf{H} \left[\mathbf{E}_v \mathbf{E}_v^\top \right]_{:,j} \\
&- 2\mathbf{H}_{:,j}^\top \left[\theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top + (\theta_v - 1) \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top \right) \right]_{:,j} \Big) \\
&\Leftrightarrow \\
&\min_{\mathbf{H}_{:,j}} \sum_{v=1}^{V} \beta_v^2 \Big(\mathbf{H}_{:,j}^\top \theta_v^2 \sum_{i=1}^{n_v} \left[\mathbf{E}_v \right]_{j,i} \mathbf{H}_{:,j} \\
&- 2\mathbf{H}_{:,j}^\top \left[\theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top + (\theta_v - 1) \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top \right) \right]_{:,j} \Big) \\
&\Leftrightarrow \\
&\min_{\mathbf{H}_{:,j}} \mathbf{H}_{:,j}^\top \sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} \left[\mathbf{E}_v \right]_{j,i} \mathbf{H}_{:,j} - 2 \cdot \\
&\mathbf{H}_{:,j}^\top \left[\sum_{v=1}^{V} \beta_v^2 \theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top + (\theta_v - 1) \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top \right) \right]_{:,j}.
\end{aligned} \tag{45}
$$

Together with the strategies for dealing with $\mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top$ during optimizing $\mathbf{C}_v$ and for dealing with $\mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top$ during optimizing $\mathbf{Q}_v$, we further have

$$
\begin{aligned}
&\min_{\mathbf{H}} \sum_{v=1}^{V} \beta_v^2 \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \\
&\Leftrightarrow \\
&\min_{\mathbf{H}_{:,j}} \mathbf{H}_{:,j}^\top \sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} \left[\mathbf{E}_v \right]_{j,i} \mathbf{H}_{:,j} - 2 \cdot \mathbf{H}_{:,j}^\top \cdot \\
&\left[\sum_{v=1}^{V} \beta_v^2 \theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \left(\mathbf{D}_v \odot \mathbf{B}_v \right) + (\theta_v - 1) \mathbf{H}_v \odot \mathbf{G}_v \right) \right]_{:,j}.
\end{aligned} \tag{46}
$$

When the coefficient $\sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} \left[\mathbf{E}_v \right]_{j,i} \neq 0$, we can simplify (46) as

$$
\min_{\mathbf{H}_{:,j}} \left\| \mathbf{H}_{:,j} - \frac{\left[\sum_{v=1}^{V} \beta_v^2 \theta_v \mathbf{Q}_v^\top \left(\widehat{\mathbf{J}}_v + \widehat{\mathbf{D}}_v \right) \right]_{:,j}}{\sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} \left[\mathbf{E}_v \right]_{j,i}} \right\|_F^2, \tag{47}
$$

where $\widehat{\mathbf{J}}_v = \mathbf{C}_v^\top \left(\mathbf{D}_v \odot \mathbf{B}_v \right)$ and $\widehat{\mathbf{D}}_v = (\theta_v - 1) \mathbf{H}_v \odot \mathbf{G}_v$. To minimize it under the constraints $\mathbf{H}_{:,j}^\top \mathbf{1}_m = 1$ and $\mathbf{H}_{:,j} \geq 0$, we formulate its Lagrange function as

$$
\mathcal{L} = \frac{1}{2} \left\| \mathbf{H}_{:,j} - \mathbf{s}_j \right\|_F^2 - \mathbf{H}_{:,j}^\top \boldsymbol{\phi} - \sigma \left(\mathbf{H}_{:,j}^\top \mathbf{1}_m - 1 \right),
$$

where $\mathbf{s}_j = \frac{\left[\sum_{v=1}^{V} \beta_v^2 \theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top (\mathbf{D}_v \odot \mathbf{B}_v) + (\theta_v - 1) \mathbf{H}_v \odot \mathbf{G}_v \right) \right]_{:,j}}{\sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i}}$, $\boldsymbol{\phi}$ and σ are Lagrange multipliers.
Then, we have

$$
\mathbf{H}_{:,j} \odot \boldsymbol{\phi} = \mathbf{0} \text{ and } \mathbf{H}_{:,j} = \mathbf{s}_j + \boldsymbol{\phi} + \sigma \mathbf{1}_m.
$$

Further, in conjunction with the column sum constraint, we can derive

$$
\mathbf{s}_j^\top \mathbf{1} + m\sigma = 1.
$$

Therefore, we have

$$
\mathbf{H}_{:,j} = \max \left(\frac{1 - \mathbf{1}^\top \mathbf{s}_j}{m} \mathbf{1} + \mathbf{s}_j, 0 \right).
$$

When $\sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} \left[\mathbf{E}_v \right]_{j,i}$ equals to 0, (46) degenerates into a linear programming problem, that is,

$$
\begin{aligned}
&\min_{\mathbf{H}_{:,j}} -\mathbf{H}_{:,j}^\top \Big[\sum_{v=1}^{V} \beta_v^2 \theta_v \mathbf{Q}_v^\top \left(\mathbf{C}_v^\top \left(\mathbf{D}_v \odot \mathbf{B}_v \right) + \right. \\
&\qquad\qquad \left. (\theta_v - 1) \mathbf{H}_v \odot \mathbf{G}_v \right) \Big]_{:,j} \\
&\qquad \text{s.t. } \mathbf{H}_{:,j}^\top \mathbf{1}_m = 1, \mathbf{H}_{:,j} \geq 0,
\end{aligned} \tag{48}
$$

which can be easily addressed by existing software.

When $\sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i} \neq 0$, (46) has a closed-form solution. It only needs to construct $\mathbf{s}_j$, taking $\mathcal{O}(md_v n + m^2 n)$ computing overhead. When $\sum_{v=1}^{V} \beta_v^2 \theta_v^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i} = 0$, it needs $\mathcal{O}(m^3 n)$ overhead via linear programming to update $\mathbf{H}$. Therefore, the computational complexity of updating $\mathbf{H}$ is $\mathcal{O}(n)$.

1.4. $\mathbf{H}_v$ Optimization

With fixed $\mathbf{C}_v$, $\mathbf{Q}_v$, $\mathbf{H}$, θ_v and β_v, the objective function (5) is equivalent to

$$\min_{\mathbf{H}_v} \sum_{v=1}^{V} \beta_v^2 \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \quad \text{s.t. } \mathbf{H}_v^\top \mathbf{1}_m = \mathbf{1}_n, \mathbf{H}_v \geq 0. \tag{49}$$

Based on the view independence, we have

$$\min_{\mathbf{H}_v} \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \quad \text{s.t. } \mathbf{H}_v^\top \mathbf{1}_m = \mathbf{1}_n, \mathbf{H}_v \geq 0. \tag{50}$$

Transforming F-norm into the trace form and deleting unrelated items, we can derive

$$\begin{aligned}
&\min_{\mathbf{H}_v} \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \\
&\Leftrightarrow \\
&\min_{\mathbf{H}_v} \operatorname{Tr} \Big((\theta_v - 1)^2 \mathbf{E}_v^\top \mathbf{H}_v^\top \mathbf{C}_v^\top \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v + 2(\theta_v - 1) \cdot \\
&\qquad \mathbf{E}_v^\top \mathbf{H}_v^\top \mathbf{C}_v^\top \left(\mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v \right) \Big) \\
&\Leftrightarrow \\
&\min_{\mathbf{H}_v} \operatorname{Tr} \Big((\theta_v - 1)^2 \mathbf{H}_v^\top \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top + 2(\theta_v - 1) \cdot \\
&\qquad \mathbf{H}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top - \theta_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top \right) \Big).
\end{aligned} \tag{51}$$

In conjunction with the column sum constraint and the unary encoding characteristic of $\mathbf{E}_v$, we further have

$$\begin{aligned}
&\min_{\mathbf{H}_v} \operatorname{Tr} \Big((\theta_v - 1)^2 \mathbf{H}_v^\top \mathbf{H}_v \mathbf{E}_v \mathbf{E}_v^\top + 2(\theta_v - 1) \cdot \\
&\qquad \mathbf{H}_v^\top \left(\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top - \theta_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top \right) \Big) \\
&\Leftrightarrow \\
&\min_{[\mathbf{H}_v]_{:,j}} [\mathbf{H}_v]_{:,j}^\top (\theta_v - 1)^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i} [\mathbf{H}_v]_{:,j} + 2(\theta_v - 1) \cdot \\
&\qquad [\mathbf{H}_v]_{:,j}^\top \left[\mathbf{C}_v^\top \mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top - \theta_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top \right]_{:,j}.
\end{aligned} \tag{52}$$

It involves constructing $\mathbf{Q}_v \mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top$, which takes $\mathcal{O}(m^2 n + mn^2)$ computing overhead. To reduce it, with the strategy tackling the matrix $\mathbf{D}_v \mathbf{E}_v \mathbf{E}_v^\top$ in optimizing $\mathbf{C}_v$, we can equivalently transform $\mathbf{H} \mathbf{E}_v \mathbf{E}_v^\top$ as $\mathbf{H} \odot \mathbf{G}_v$, where $\mathbf{G}_v$ has been constructed in optimizing $\mathbf{Q}_v$. Accordingly, we can obtain

$$\begin{aligned}
&\min_{\mathbf{H}_v} \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \\
&\Leftrightarrow \\
&\min_{[\mathbf{H}_v]_{:,j}} [\mathbf{H}_v]_{:,j}^\top (\theta_v - 1)^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i} [\mathbf{H}_v]_{:,j} + 2(\theta_v - 1) \cdot \\
&\qquad [\mathbf{H}_v]_{:,j}^\top \left[\mathbf{C}_v^\top (\mathbf{D}_v \odot \mathbf{B}_v) - \theta_v \mathbf{Q}_v (\mathbf{H} \odot \mathbf{G}_v) \right]_{:,j}.
\end{aligned} \tag{53}$$

When $(\theta_v - 1)^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i}$ is non-zero, it can be further simplified as

$$\min_{[\mathbf{H}_v]_{:,j}} \left\| [\mathbf{H}_v]_{:,j} + \frac{\left[\mathbf{C}_v^\top (\mathbf{D}_v \odot \mathbf{B}_v) - \theta_v \mathbf{Q}_v (\mathbf{H} \odot \mathbf{G}_v) \right]_{:,j}}{(\theta_v - 1) \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i}} \right\|_F^2. \tag{54}$$

In conjunction with the strategy handling with (47), we can derive

$$[\mathbf{H}_v]_{:,j} = \max \left(\frac{1 + \mathbf{1}^\top \mathbf{t}_j^v}{m} \mathbf{1} - \mathbf{t}_j^v, 0 \right), \tag{55}$$

where the vector $\mathbf{t}_j^v = \frac{\left[\mathbf{C}_v^\top (\mathbf{D}_v \odot \mathbf{B}_v) - \theta_v \mathbf{Q}_v (\mathbf{H} \odot \mathbf{G}_v) \right]_{:,j}}{(\theta_v - 1) \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i}}$.

When $(\theta_v - 1)^2 \sum_{i=1}^{n_v} [\mathbf{E}_v]_{j,i}$ is zero, (53) becomes a linear programming problem, and can be solved through off-the-shelf packages.

The construction of $\mathbf{C}_v^\top (\mathbf{D}_v \odot \mathbf{B}_v)$ and $\mathbf{Q}_v (\mathbf{H} \odot \mathbf{G}_v)$ takes $\mathcal{O}(md_v n)$ and $\mathcal{O}(m^2 n)$ respectively. The linear programming takes $\mathcal{O}(m^3 n)$ overhead. Therefore, updating $\mathbf{H}_v$ requires $\mathcal{O}(n)$ computational complexity.

1.5. θ_v Optimization

With fixed $\mathbf{C}_v$, $\mathbf{Q}_v$, $\mathbf{H}$, $\mathbf{H}_v$ and β_v, the objective function (5) transforms into

$$\min_{\theta_v} \sum_{v=1}^{V} \beta_v^2 \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \quad \text{s.t. } 0 \leq \theta_v \leq 1. \tag{56}$$

According to the view independence, it is further reducible to the following problem,

$$\min_{\theta_v} \left\| \mathbf{D}_v \mathbf{E}_v - \theta_v \mathbf{C}_v \mathbf{Q}_v \mathbf{H} \mathbf{E}_v + (\theta_v - 1) \mathbf{C}_v \mathbf{H}_v \mathbf{E}_v \right\|_F^2 \quad \text{s.t. } 0 \leq \theta_v \leq 1. \tag{57}$$

Then, by virtue of trace expanding operation, we can derive

$$\begin{aligned}
&\min_{\theta_v} \|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2 \\
&\Leftrightarrow \\
&\min_{\theta_v} \|\mathbf{D}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{H}_v\mathbf{E}_v + \theta_v\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\|_F^2 \\
&\Leftrightarrow \\
&\min_{\theta_v} \mathrm{Tr}\Big(\theta_v^2\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\cdot \\
&\qquad \left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)^\top + \\
&\quad 2\theta_v\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\left(\mathbf{D}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right)^\top\Big).
\end{aligned} \tag{58}$$

The optimal solution of this unconstrained optimization task is

$$\theta_v = \frac{\mathrm{Tr}\left(\left(\widehat{\mathbf{L}}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\left(\widehat{\mathbf{L}}_v - \mathbf{D}_v\mathbf{E}_v\right)^\top\right)}{\mathrm{Tr}\left(\left(\widehat{\mathbf{L}}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\left(\widehat{\mathbf{L}}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)^\top\right)},$$

where $\widehat{\mathbf{L}}_v = \mathbf{C}_v\mathbf{H}_v\mathbf{E}_v$.

It needs to construct the items $(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v)$ and $(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{D}_v\mathbf{E}_v)$, which takes $\mathcal{O}(d_vm^2 + d_vmn + d_vnn_v)$ and $\mathcal{O}(d_vmn + d_vnn_v)$ computing overhead, respectively. Evidently, it is close to the quadratic cost with respect to the sample size n. To decrease the overhead, we observe

$$\begin{aligned}
&\mathrm{Tr}\left(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{D}_v\mathbf{E}_v\right)^\top\right) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\cdot \\
&\qquad\qquad \left(\mathbf{E}_v^\top\mathbf{H}_v^\top\mathbf{C}_v^\top - \mathbf{E}_v^\top\mathbf{D}_v^\top\right)\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top\mathbf{C}_v^\top - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top\mathbf{C}_v^\top \\
&\qquad - \mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{D}_v^\top + \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{D}_v^\top\Big).
\end{aligned} \tag{59}$$

Combined with the trace cycle operation, we can further derive

$$\begin{aligned}
&\mathrm{Tr}\left(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{D}_v\mathbf{E}_v\right)^\top\right) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top - \mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top \\
&\qquad - \mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{D}_v^\top + \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{D}_v^\top\Big).
\end{aligned} \tag{60}$$

For the item $\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v$, we have

$$\begin{aligned}
&\mathrm{Tr}\Big(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\cdot \\
&\qquad\qquad \left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)^\top\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\cdot \\
&\qquad\qquad \left(\mathbf{E}_v^\top\mathbf{H}_v^\top\mathbf{C}_v^\top - \mathbf{E}_v^\top\mathbf{H}^\top\mathbf{Q}_v^\top\mathbf{C}_v^\top\right)\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top\mathbf{C}_v^\top - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top\mathbf{C}_v^\top - \\
&\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}^\top\mathbf{Q}_v^\top\mathbf{C}_v^\top + \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}^\top\mathbf{Q}_v^\top\mathbf{C}_v^\top\Big).
\end{aligned} \tag{61}$$

Therefore, we can obtain

$$\begin{aligned}
&\mathrm{Tr}\Big(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\cdot \\
&\qquad\qquad \left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)^\top\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top - \mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}_v^\top \\
&\qquad - \mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}^\top\mathbf{Q}_v^\top + \mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top\mathbf{H}^\top\mathbf{Q}_v^\top\Big).
\end{aligned} \tag{62}$$

Combined with the scheme tackling $\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top$, we have

$$\begin{aligned}
&\mathrm{Tr}\left(\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\right)\left(\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{D}_v\mathbf{E}_v\right)^\top\right) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\left(\mathbf{H}_v \odot \mathbf{G}_v\right)\mathbf{H}_v^\top - \mathbf{Q}_v\mathbf{H}\left(\mathbf{H}_v \odot \mathbf{G}_v\right)^\top \\
&\qquad - \mathbf{C}_v\mathbf{H}_v\left(\mathbf{D}_v \odot \mathbf{B}_v\right)^\top + \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\left(\mathbf{D}_v \odot \mathbf{B}_v\right)^\top\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\left(\mathbf{H}_v \odot \mathbf{G}_v\right)\mathbf{H}_v^\top - \left(\mathbf{H}_v \odot \mathbf{G}_v\right)\mathbf{H}^\top\mathbf{Q}_v^\top \\
&\qquad - \mathbf{C}_v\mathbf{H}_v\left(\mathbf{D}_v \odot \mathbf{B}_v\right)^\top + \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\left(\mathbf{D}_v \odot \mathbf{B}_v\right)^\top\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\left(\mathbf{H}_v \odot \mathbf{G}_v\right)\left(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H}\right)^\top \\
&\qquad - \mathbf{C}_v\left(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H}\right)\left(\mathbf{D}_v \odot \mathbf{B}_v\right)^\top\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\Big(\left(\mathbf{H}_v \odot \mathbf{G}_v\right)\left(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H}\right)^\top \\
&\qquad - \mathbf{C}_v^\top\left(\mathbf{D}_v \odot \mathbf{B}_v\right)\left(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H}\right)^\top\Big) \\
&\Leftrightarrow \\
&\mathrm{Tr}\left(\left(\mathbf{H}_v \odot \mathbf{G}_v - \mathbf{C}_v^\top\left(\mathbf{D}_v \odot \mathbf{B}_v\right)\right)\left(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H}\right)^\top\right)
\end{aligned} \tag{63}$$

and

$$\operatorname{Tr}\Big((\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v)\cdot (\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v - \mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v)^\top \Big)$$
$$\Leftrightarrow$$
$$\operatorname{Tr}\Big((\mathbf{H}_v \odot \mathbf{G}_v)\,\mathbf{H}_v^\top - \mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v)\,\mathbf{H}_v^\top - (\mathbf{H}_v \odot \mathbf{G}_v)\,\mathbf{H}^\top\mathbf{Q}_v^\top + \mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v)\,\mathbf{H}^\top\mathbf{Q}_v^\top \Big)$$
$$\Leftrightarrow$$
$$\operatorname{Tr}\Big((\mathbf{H}_v \odot \mathbf{G}_v - \mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v))\,(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H})^\top \Big). \tag{64}$$

Constructing $(\mathbf{H}_v \odot \mathbf{G}_v - \mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v))\,(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H})^\top$ and $(\mathbf{H}_v \odot \mathbf{G}_v - \mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v))\,(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H})^\top$ takes $\mathcal{O}(md_vn + m^2n)$ and $\mathcal{O}(m^2n)$ computing overhead respectively, both of which are linear to the sample number. Combined with the constraint conditions on θ_v, therefore, we have

$$\theta_v = \operatorname{clip}\left(\frac{\operatorname{Tr}\Big(\Big(\widehat{\mathbf{K}}_v - \mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v)\Big)\widehat{\mathbf{E}}_v^\top\Big)}{\operatorname{Tr}\Big(\Big(\widehat{\mathbf{K}}_v - \mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v)\Big)\widehat{\mathbf{E}}_v^\top\Big)}, 0, 1 \right),$$

where $\widehat{\mathbf{K}}_v = \mathbf{H}_v \odot \mathbf{G}_v$ and $\widehat{\mathbf{E}}_v = \mathbf{H}_v - \mathbf{Q}_v\mathbf{H}$.

1.6. β_v Optimization

With fixed $\mathbf{C}_v$, $\mathbf{Q}_v$, $\mathbf{H}$, $\mathbf{H}_v$ and θ_v, the objective function (5) reduces to

$$\min_{\beta_v} \sum_{v=1}^{V} \beta_v^2 \left\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\right\|_F^2 \quad \text{s.t. } \boldsymbol{\beta} \geq 0, \boldsymbol{\beta}^\top\mathbf{1} = 1. \tag{65}$$

Since $\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2$ is a constant about β_v, with Cauchy inequality, we can derive

$$\beta_v = \frac{\frac{1}{\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2}}{\sum_{v=1}^{V}\frac{1}{\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2}}. \tag{66}$$

Calculating $\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2$ needs $\mathcal{O}(d_vnn_v + d_vm^2 + d_vmn)$ computing overhead, which approaches $\mathcal{O}(n^2)$. To reduce this overhead, we notice that each column of the matrix $\mathbf{E}_v$ has a single 1 and other entries are equal to 0. On the basis of this, we can derive that $\|\mathbf{D}_v\mathbf{E}_v\|_F^2$ is equal to $\left\|\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top\right\|_F^2$. Accordingly, we have the following equation holds true,

$$\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2 = \left\|\mathbf{D}_v\mathbf{E}_v\mathbf{E}_v^\top - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v\mathbf{E}_v^\top + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\mathbf{E}_v^\top\right\|_F^2. \tag{67}$$

Further, in conjunction with the strategies tackling $\mathbf{D}_v$, $\mathbf{H}$, $\mathbf{H}_v$ and $\mathbf{E}_v\mathbf{E}_v^\top$, we have

$$\|\mathbf{D}_v\mathbf{E}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\mathbf{H}\mathbf{E}_v + (\theta_v - 1)\mathbf{C}_v\mathbf{H}_v\mathbf{E}_v\|_F^2 = \|\mathbf{D}_v\odot\mathbf{B}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v) + (\theta_v - 1)\mathbf{C}_v\,(\mathbf{H}_v\odot\mathbf{G}_v)\|_F^2. \tag{68}$$

As a result, we can obtain

$$\beta_v = \frac{\frac{1}{\|\mathbf{D}_v\odot\mathbf{B}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v(\mathbf{H}\odot\mathbf{G}_v) + (\theta_v - 1)\mathbf{C}_v(\mathbf{H}_v\odot\mathbf{G}_v)\|_F^2}}{\sum_{v=1}^{V}\frac{1}{\|\mathbf{D}_v\odot\mathbf{B}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v(\mathbf{H}\odot\mathbf{G}_v) + (\theta_v - 1)\mathbf{C}_v(\mathbf{H}_v\odot\mathbf{G}_v)\|_F^2}}.$$

The building of $\left\|\mathbf{D}_v\odot\mathbf{B}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v) + \widehat{\mathbf{F}}_v\right\|_F^2$ where the matrix $\widehat{\mathbf{F}}_v = (\theta_v - 1)\mathbf{C}_v\,(\mathbf{H}_v\odot\mathbf{G}_v)$ needs $\mathcal{O}(d_vm^2 + d_vmn)$ computing overhead. Additionally, the data dimension d_v is irrelevant to n, and m is largely smaller than n. Therefore, updating β_v takes $\mathcal{O}(dmn)$ overhead where d is $\sum_{v=1}^{V} d_v$. That is, the computational complexity about β_v is $\mathcal{O}(n)$.

2. Complexity Discussion

2.1. Time Complexity

Optimizing the variables $\mathbf{C}_v$, $\mathbf{Q}_v$, $\mathbf{H}$, $\mathbf{H}_v$, θ_v and β_v needs $\mathcal{O}(m^2n + d_vnm + d_vm^2)$, $\mathcal{O}(md_vn + m^2n)$, $\mathcal{O}(md_vn + m^3n)$, $\mathcal{O}(md_vn + m^3n)$, $\mathcal{O}(md_vn + m^2n)$ and $\mathcal{O}(dmn)$ computing overhead respectively. Consequently, the time complexity is $\mathcal{O}(dm^2 + m^3n + dmn)$. Generally, $m \ll n$ and d is a constant. Hence, our time complexity is $\mathcal{O}(n)$.

2.2. Space Complexity

The space complexity consists of the memory cost of updating $\mathbf{C}_v$, $\mathbf{Q}_v$, $\mathbf{H}$, $\mathbf{H}_v$, θ_v and β_v. During updating $\mathbf{C}_v$, the construction of $\mathbf{D}_v\odot\mathbf{B}_v\,(\theta_v\mathbf{Q}_v\mathbf{H} - (\theta_v - 1)\mathbf{H}_v)^\top$ needs $\mathcal{O}(mn + d_vn + d_vm)$ memory cost. During updating $\mathbf{Q}_v$, the construction of $\left(\mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v) + (\theta_v - 1)\,(\mathbf{H}_v\odot\mathbf{G}_v)\right)\mathbf{H}^\top$ needs $\mathcal{O}(mn + m^2)$. During updating $\mathbf{H}$, since $\mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v)$ and $(\mathbf{H}_v\odot\mathbf{G}_v)$ have been established in the updating of $\mathbf{Q}_v$, it only needs $\mathcal{O}(Vmn)$ space cost to construct $\sum_{v=1}^{V}\beta_v^2\theta_v\mathbf{Q}_v^\top\left(\mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v) + (\theta_v - 1)\mathbf{H}_v\odot\mathbf{G}_v\right)$. During updating $\mathbf{H}_v$, since $\mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v)$ has been established, it only needs $\mathcal{O}(mn)$ to construct $\mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v) - \theta_v\mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v)$. During updating θ_v, it only needs $\mathcal{O}(mn + m^2)$ and $\mathcal{O}(m^2)$ to construct $(\mathbf{H}_v\odot\mathbf{G}_v - \mathbf{C}_v^\top(\mathbf{D}_v\odot\mathbf{B}_v))\,(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H})^\top$ and $(\mathbf{H}_v\odot\mathbf{G}_v - \mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v))\,(\mathbf{H}_v - \mathbf{Q}_v\mathbf{H})^\top$. During updating β_v, since $\mathbf{D}_v\odot\mathbf{B}_v$, $\mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v)$ and $\mathbf{H}_v\odot\mathbf{G}_v$ have been established, it only needs $\mathcal{O}(nd)$ memory cost to construct the item $\left\|\mathbf{D}_v\odot\mathbf{B}_v - \theta_v\mathbf{C}_v\mathbf{Q}_v\,(\mathbf{H}\odot\mathbf{G}_v) + \widehat{\mathbf{F}}_v\right\|_F^2$ where $\widehat{\mathbf{F}}_v = (\theta_v - 1)\mathbf{C}_v\,(\mathbf{H}_v\odot\mathbf{G}_v)$. Thus, it totally takes

$\mathcal{O}(d_v n + d_v m + m^2 + Vmn + nd)$ memory cost. Combined with the fact that $m \ll n$ and that d and V are constants, we have that our space complexity is $\mathcal{O}(Vmn + nd)$. Evidently, it is $\mathcal{O}(n)$.

3. Experimental Setting

We initialize $\mathbf{C}_v$ with an orthogonal matrix and $\mathbf{Q}_v$ with a randomly-shuffled identity matrix. For $\mathbf{H}$ and $\mathbf{H}_v$, we initialize them with randomly-shuffled one-hot vectors. For θ_v and β_v, we initialize them with $1/2$ and $1/V$ respectively. The stopping condition is set as $(f_{t-1} - f_t)/f_{t-1} \leq 1e-4$ where f_t denotes the objective value at the t-th iteration. We perform twenty runs for each algorithm, and report the average clustering results. Three metrics, accuracy (ACC), purity (PUR), fscore (FSC) are employed to evaluate the clustering quality.

4. Introduction to Comparison Algorithms

- HCMSC: It leverages tensor factorization strategy to exploit correlations among samples and views, and constructs a hypergraph structure to regularize the inference of incomplete views and the learning of cluster subspaces.
- LSIVC: It treats sparse representation learning and graph embedding as a single term through localized constraints, and utilizes an information entropy based balancing strategy to actively calibrate respective view contributions.
- LRIVC: It combines graph completion and spectral embedding to restore missing similarity structures, and enforces a connectivity constraint to exploit intra-view relationship and meanwhile maintain semantic consistency.
- AGIMC: It integrates within-view preservation and cross-view inference to achieve a joint graph completion, and utilizes a weighted multi-graph geometric structure to infer missing features and establish unified representation.
- IVCBG: It introduces view-related projectors to render view data with diverse dimensions into a latent shared space, and constructs a unified graph via a set of unified anchors to build missing connections and perform fusion.
- PIMVC: It alleviates information imbalance via extracting features in low-dimensional subspace, and employs graph regularization and uncorrelated constraints to maintain data geometric structure and avoid trivial solutions.
- GRIMC: It incorporates tensor nuclear norm to decouple biased errors from partial subgraphs and avoid feature restore procedure, and utilizes corss-view learning mechanism to extract local structure and refine graph structure.
- UIMC: It decouples self-representation matrices using a matching graph to infer incomplete features, and introduces a tensor decomposition mode via an exploratory parameter-adjusting strategy to exploit view correlations.
- OSIMC: It integrates base partition filling and cluster label learning to improve mutual guidance between feature imputation and clustering objectives, and introduces common partition learning to construct efficient late-fusion.
- USETL: It unifies spectral embedding learning and low-rank learning to decrease redundant information and improve view-information consistency, and constructs an incomplete-instance-aware strategy to generate similarity.
- LBIVC: It introduces probability-formulated representation learning and graph regularization to eliminate post-processing requirement, and devises a balancing constraint to alleviate over-concentration on certain samples.
- HCLGL: It constructs a confidence graph through neighbor strategy to encode group-wise local geometry, and employs a Laplacian rank constraint to assist in generating the shared hypergraph with block-diagonal structure.

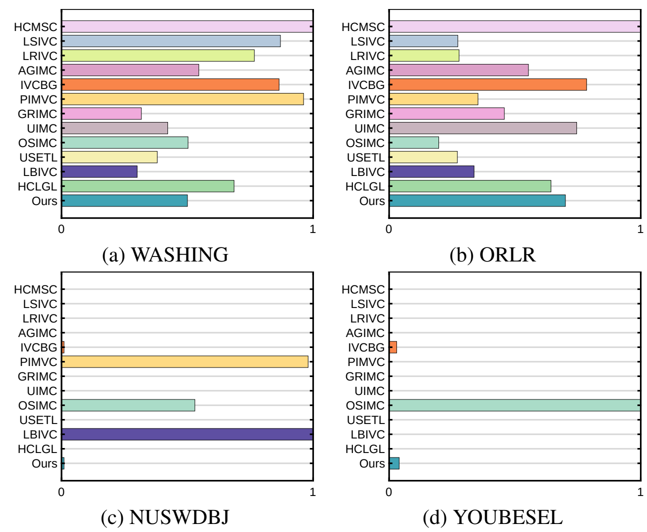
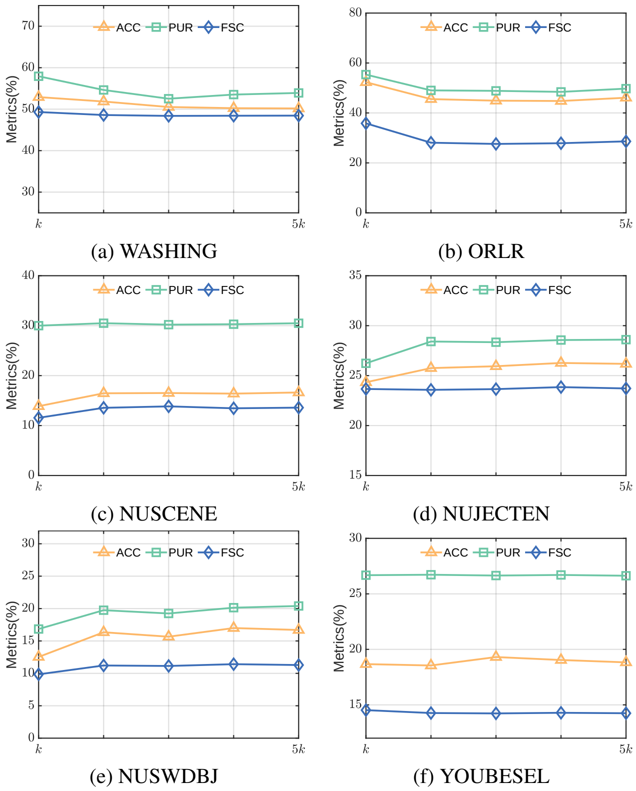


Figure 5. Memory Overhead Comparison on Other Four Datasets

5. More Observations for Performance Table 1

- OSIMC achieves two marginally superior clustering results on WASHING, which could be attributed to its unified framework that jointly performs data imputation and cluster label learning. The consensus partition derived from clustering, in turn, provides guidance for the feature recovery process in a synergistic manner.
- For LSIVC, it outperforms us in PUR on NUSCENE at the 60% missing ratio, which is possibly owing to its effective integration of graph embedding learning with consistency learning, coupled with a localized sparse regularizer that extracts structured representations to mitigate the imbalance factors induced by view incompleteness.
- On NUSCENE and NUJECTEN, LRIVC makes better results in PUR, possibly because it builds view relationship via adaptive graph embedding to capture the underlying data manifold, and integrates related information within



(a) WASHING (b) ORLR

(c) NUSCENE (d) NUJECTEN

(e) NUSWDBJ (f) YOUBESEL

Figure 6. Anchor Quantity Influence

views into a consistent global structure to leverage low-rank characteristics across views.

- The performance of IVCBG on ORLR and YOUBESEL demonstrates advantages, the underlying reasons of which could lie in its strategy of leveraging random walks to construct a common similarity with probability transition property while utilizing specialized projectors to bridge the dimensional gap between views.
- PIMVC delivers some favorable results on NUSWDBJ, possible reasons of which are that it employs a graph regularizr to exploit structure features in refined subspace, and leverages the mapping transformation between original space and potential subspace to dynamically balance view contributions.

6. More Observations for Running Time Fig. 2

- Methods such as AGIMC, LSIVC, HCLGL, etc, are more computationally intensive than our approach. This is primarily due to their reliance on constructing a self-expression matrix. In contrast, our approach efficiently builds a compact anchor-based affinity, resulting in significantly lower computational time.
- GRIMC, USETL, UIMC, etc, exhibit longer runtime than us. This performance disparity is primarily attributable to their incorporation of tensor constraints, which invariably incurs a high computational overhead due to the complexity of the associated tensor operations though powerful for data structure preservation.
- IVCBG, PIMVC and LBIVC exhibit a slight speed advantage over us, which is likely owing to the low-dimensional projectors. However, due to IVCBG failing to account for view-wise dissimilarities, PIMVC being constrained by its use of simplistic linear mappings, and LBIVC suffering from a disjointed handling of geometric structure, they typically produce inferior results.

7. Other Memory Overhead Comparison

Fig. 5 presents the memory overhead comparison on the rest of datasets. As shown, our overhead is still comparable across most benchmarks, especially on large-scale datasets NUSWDBJ and YOUBESEL where we receive a notably lower memory requirement. On ORLR, it seems that our overhead appears somewhat higher, possibly because the inherent design of our view-exclusive artifact learning and topology harmonization components necessitate additional storage space.

Table 4. Ablation Study

Ablation	WASHING			ORLR			NUSCENE		
	ACC	PUR	FSC	ACC	PUR	FSC	ACC	PUR	FSC
OVW	49.83	53.04	48.35	48.70	52.29	32.04	13.35	29.15	11.17
OAL	39.17	57.28	37.56	41.53	44.20	12.83	11.76	29.21	9.24
Ours	**52.91**	**57.93**	**49.31**	**52.31**	**55.36**	**35.84**	**13.87**	**30.00**	**11.55**
	NUJECTEN			NUSWDBJ			YOUBESEL		
OVW	24.14	26.16	23.36	11.81	16.83	9.69	18.42	**26.69**	14.39
OAL	20.79	24.89	20.27	11.92	**16.94**	9.61	16.84	25.66	13.72
Ours	**24.33**	**26.23**	**23.66**	**12.53**	16.84	**9.85**	**18.67**	26.68	**14.53**

8. Other Ablations

In addition to view-exclusive artifact learning and topology harmonization, we also organize the ablation study on the employed view weighting (VW) and anchor learning (AL) strategies, and the results are presented in Table 4 where OVW and OAL represent no view weighting and anchor learning respectively. Evidently, our VW and AL are functional, and facilitate performance boosting.

9. Influence of Anchor Quantity

In this study, the anchor quantity is designed to be the same as the cluster quantity k. It can also be set to other values. To investigate the influence of anchor quantity on the model performance, we organize experiments under anchor quantity as $k \sim 5k$ respectively. The comparative results are summarized in Fig. 6. As observed, the clustering metrics exhibit only minor fluctuations, which illustrates that our proposed model is relatively robust with respect to the anchor quantity.

10. Stability

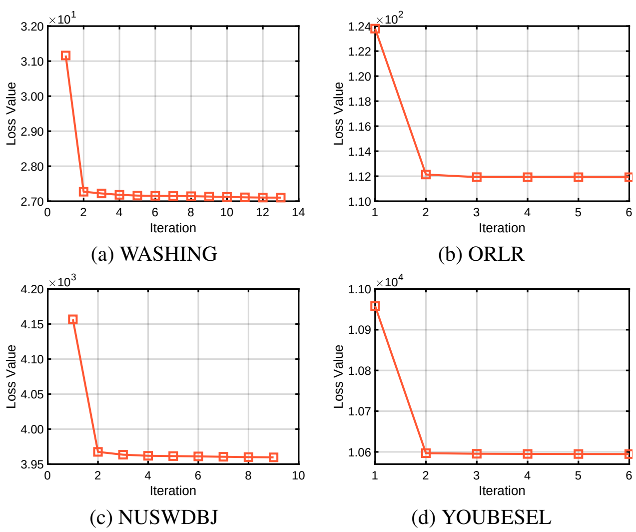
Besides the mean values, we report the standard deviation of the clustering results to assess the stability of our model, as summarized in Table 5. The low variance observed across experiments indicates that our method is highly consistent and yields stable clustering results.

Table 5. Standard Deviation

Dataset	30%			60%			80%		
	ACC	PUR	FSC	ACC	PUR	FSC	ACC	PUR	FSC
WASHING	1.45	1.53	0.46	1.38	2.18	0.42	0.29	0.57	0.04
ORLR	1.45	1.22	1.82	2.13	1.90	1.78	2.21	1.91	2.20
NUSCENE	0.62	0.22	0.52	0.58	0.18	0.37	0.55	0.28	0.40
NUJECTEN	0.25	0.14	0.19	0.65	0.86	0.13	0.18	0.29	0.15
NUSWDBJ	0.69	0.66	0.32	0.50	0.31	0.11	0.54	0.38	0.13
YOUBESEL	0.35	0.03	0.11	0.22	0.02	0.10	0.26	0.01	0.14

11. Other Convergence Results

Fig. 7 presents the loss value evolution on the rest of four datasets. As observed, the loss curves exhibit a consistent monotonic downward trend and eventual stabilization, which empirically confirms the convergence of our proposed approach.



(a) WASHING (b) ORLR

(c) NUSWDBJ (d) YOUBESEL

Figure 7. Loss Value Evolution on Other Four Datasets