LOW-RANK RECOVERY OF HIGH-ORDER NEIGHBOR STRUCTURE FOR SCALABLE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Anchor-based multi-view clustering has attracted considerable attention for its scalability. However, existing methods face three limitations: (1) direct sample–anchor graphs insufficiently capture high-order neighbor connectivity; (2) conventional tensor nuclear norm regularization may over-shrink dominant singular values; (3) heterogeneous relationships among view pairs are often not explicitly modeled. To address these limitations, we propose **H**igh-**O**rder **N**eighbor **R**ecovery (HONR), a scalable framework for recovering high-order neighbor structure through low-rank tensor learning. Specifically, HONR uses multi-step random walks to capture indirect sample–anchor connections and aggregates transitions across orders and views into a compact tensor. For a more accurate approximation of tensor rank, we introduce an arctangent-based nonconvex approximation that alleviates the excessive shrinkage of dominant singular values. Furthermore, a view-relation regularizer based on order-wise transition similarities explicitly models pairwise view relationships and promotes consistency between structurally similar views. An alternating optimization algorithm is developed, with computational complexity linear in the number of samples when the numbers of anchors and views are fixed. Experiments against eight baselines on seven benchmark datasets demonstrate the effectiveness of the proposed HONR.

1 INTRODUCTION

Real-world objects are often described by multiple modalities or heterogeneous feature sets, giving rise to multi-view data. Multi-view clustering (MVC) aims to exploit the complementary and consistent information across these views to uncover the underlying clustering structure. Existing MVC approaches can be broadly categorized into graph-based methods (Huang et al., 2022; Liang et al., 2024), low-rank and subspace-based methods (You et al., 2023; Fu et al., 2023), and deep learning-based methods (Li et al., 2025). Despite their promising performance, many graph-based and subspace-based approaches require constructing and processing sample-level similarity or representation matrices, leading to substantial computational and storage costs as the number of samples increases.

To address this scalability issue, anchor-based MVC represents samples using a small set of representative anchors and constructs compact sample–anchor graphs instead of full sample–sample graphs (Sakai & Imiya, 2009; Zhang & Lu, 2016; Nie et al., 2017; Ji & Feng, 2023; Li et al., 2022). With n samples and $m \ll n$ anchors, this paradigm reduces the graph size from $n \times n$ to $n \times m$. This substantially lowers storage costs and facilitates large-scale MVC. For instance, Kang et al. (2019) proposed LMVSC to learn and integrate compact view-specific anchor graphs for efficient spectral clustering, achieving linear complexity with respect to the number of samples. Similarly, Li et al. (2022) constructed view-specific bipartite graphs and employed a self-weighted graph fusion strategy for scalable and parameter-free MVC. By jointly performing anchor selection and subspace graph construction, Wang et al. (2022) developed a parameter-free framework for large-scale multi-view subspace clustering under consensus anchor guidance. To address cross-view heterogeneity, Ou et al. (2023) projected different views into a common subspace through hierarchical feature descent and subsequently performed anchor-based subspace clustering with linear computational complexity.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Recent studies have incorporated tensor learning into anchor-based MVC to capture cross-view correlations. Zhao et al. (2024) integrated feature selection, sparse anchor-graph refinement, and tensor-based multi-view learning to suppress noisy anchors and redundant connections, while directly producing clustering labels without additional post-processing. Du et al. (2025) proposed WTCAGL, which employs a weighted tensor nuclear norm to jointly model view-specific and consistent anchor graphs that capture high-order cross-view correlations for effective multi-view clustering. Wang et al. (2026) developed CATMSC, which combines low-rank anchor tensor learning with correntropy for robust multi-view clustering. Long et al. (2024) proposed S^2MVTC, which jointly captures within-view smoothness and cross-view semantic correlations through tensor-based embedding feature learning for scalable multi-view clustering. Nevertheless, three challenges motivate further investigation. First, direct sample–anchor connections primarily describe immediate relationships; exploiting cross-view tensor correlations does not by itself explicitly encode multi-step neighborhood connectivity within each view. Second, conventional tensor nuclear norm regularization penalizes singular values linearly, potentially over-shrinking dominant components that carry useful clustering information. Third, although existing methods exploit cross-view consistency or correlations, the heterogeneous relationships among different view pairs are not always explicitly modeled, potentially limiting the effective use of inter-view structural information.

To this end, we propose High-Order Neighbor Recovery (HONR), a scalable multi-view clustering framework that recovers high-order neighbor structure through low-rank tensor learning. We first perform multi-step random walks over anchors and samples to aggregate transitions across different orders within each view. These matrices are then stacked across views to form a compact cross-order anchor transition tensor, incorporating both immediate and high-order neighborhood information. Next, we recover a clean representation through low-rank tensor learning with an arctangent-based nonconvex rank approximation, which reduces the penalization of large singular values relative to small ones. Furthermore, a view-relation graph explicitly characterizes heterogeneous pairwise relationships among views and guides cross-view structural consistency. The key contributions of this work are summarized as follows:

- We propose HONR, a scalable MVC framework that recovers high-order neighbor structure through low-rank tensor learning. Multi-step random walks construct a compact cross-order anchor transition tensor, while view-relation regularization strengthens consistency between structurally similar views. The framework scales linearly with the number of samples when other dimensions and iteration counts are fixed.
- We develop an arctangent-based nonconvex tensor rank approximation to recover the underlying low-rank structure while alleviating the shrinkage of dominant singular values. Moreover, we derive a closed-form solution for the associated subproblems, enabling efficient model optimization.
- Experiments against eight baselines on seven benchmark datasets confirm the significant effectiveness and efficiency of HONR.

2 Related Work and Preliminaries

We use bold calligraphic letters, bold uppercase letters, and bold lowercase letters (e.g., $\boldsymbol{\mathcal{A}}$, $\mathbf{A}$, and $\mathbf{a}$) to denote tensors, matrices, and vectors, respectively. For a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{\mathcal{A}}^{(k)}$ denotes its k-th frontal slice. Let $\hat{\boldsymbol{\mathcal{A}}} = \mathbf{fft}(\boldsymbol{\mathcal{A}}, [], \mathbf{3}) \in \mathbb{C}^{\mathbf{n_1} \times \mathbf{n_2} \times \mathbf{n_3}}$ denote the fast Fourier transform (FFT) along the third dimension, with the inverse fast Fourier transform (IFFT) given by $\boldsymbol{\mathcal{A}} = \mathbf{ifft}(\hat{\boldsymbol{\mathcal{A}}}, [], \mathbf{3})$. The t-SVD can be computed by applying matrix SVD to each frontal slice of $\hat{\boldsymbol{\mathcal{A}}}$, followed by the inverse Fourier transform of the resulting factors.

Definition 2.1 (Tensor Nuclear Norm, TNN). *The tensor nuclear norm of $\boldsymbol{\mathcal{A}}$ is defined as*

$$\|\boldsymbol{\mathcal{A}}\|_{\circledast} = \sum_{k=1}^{n_3} \|\hat{\boldsymbol{\mathcal{A}}}^{(k)}\|_* = \sum_{k=1}^{n_3} \sum_{i=1}^{\min(n_1,n_2)} \sigma_i\left(\hat{\boldsymbol{\mathcal{A}}}^{(k)}\right), \tag{1}$$

where $\|\cdot\|_*$ denotes the matrix nuclear norm and $\sigma_i(\cdot)$ denotes the i-th largest singular value.

2.1 Anchor-based Multi-view Clustering

To alleviate the computational burden caused by constructing full sample-to-sample affinity graphs, anchor-based methods select or generate a small number of representative data points, termed anchors or landmarks (Cai & Chen, 2015; Liu et al., 2010; Han et al., 2017), typically through random sampling (Sakai & Imiya, 2009) or k-means clustering (Kang et al., 2019). Based on these anchors, a compact bipartite graph is constructed to characterize the relationships between samples and anchors. In multi-view clustering, such methods generally construct view-specific anchor graphs to preserve the underlying geometric structures of different views. Specifically, consider a multi-view dataset $\{X^{(v)}\}_{v=1}^{V}$ with V views, where $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \ldots, x_n^{(v)}] \in \mathbb{R}^{d_v \times n}$ denotes the data matrix of the v-th view, and d_v is its feature dimension. Given the anchor matrix $U^{(v)} = [u_1^{(v)}, u_2^{(v)}, \ldots, u_m^{(v)}] \in \mathbb{R}^{d_v \times m}$, the corresponding anchor graph $M^{(v)} \in \mathbb{R}^{n \times m}$ is defined by $(M^{(v)})_{ij} = \texttt{sim}(x_i^{(v)}, u_j^{(v)})$ if $j \in \mathcal{N}_k(i)$, and 0 otherwise, where $\texttt{sim}(\cdot,\cdot)$ measures sample–anchor affinity and $\mathcal{N}_k(i)$ indexes the k nearest anchors of sample i. The view-specific anchor graphs are then integrated to obtain the final clustering result.

2.2 Low-rank Tensor recovery

Tensor learning has been widely adopted in multi-view clustering to exploit high-order correlations among multiple views. By stacking view-specific representations into a third-order tensor, low-rank tensor regularization can effectively capture shared structures underlying different views. A representative framework for low-rank tensor recovery is tensor robust principal component analysis (TRPCA), which decomposes an observed tensor into a low-rank component and a sparse error component (Lu et al., 2016). For example, Wu et al. (2019) constructed a transition probability tensor $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{n \times n \times V}$ from multiple views and recovered its underlying clean structure $\boldsymbol{\mathcal{Z}}$ through

$$\min_{\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{Z}}\|_{\circledast} + \lambda \|\boldsymbol{\mathcal{E}}\|_{2,1}, \quad \text{s.t. } \boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{Z}} + \boldsymbol{\mathcal{E}}, \tag{2}$$

where λ is a balancing parameter, $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{n \times n \times V}$ denotes the noise tensor, and $\|\cdot\|_{2,1}$ denotes the sum of ℓ_2 norms of all mode-2 fibers. In particular, each frontal slice of $\boldsymbol{\mathcal{P}}$ is constructed as a transition probability matrix of the samples in the corresponding view.

3 Proposed Method

3.1 Cross-Order Anchor Transition Tensor Construction

For anchor selection, we concatenate the features from all views and apply k-means to the resulting representation, selecting the sample nearest to each cluster center as an anchor. To construct the anchor graph $M^{(v)}$ according to section 2.1, we adopt a Gaussian similarity function and retain k nearest anchors for each sample. The Gaussian kernel bandwidth σ_v is set to the median of all sample–anchor distances in the v-th view. However, $M^{(v)}$ captures only direct sample–anchor connections and does not explicitly encode higher-order neighborhood relationships. Multi-step random walks provide a natural way to capture such relationships by propagating transition probabilities along paths in the graph (Peng et al., 2024a;b).

To this end, we first normalize the anchor graph into a sample–anchor transition probability matrix: $P^{(v)} = (D^{(v)})^{-1} M^{(v)}$, where $D^{(v)}$ is a diagonal matrix with $D_{ii}^{(v)} = \sum_{j=1}^{m} M_{ij}^{(v)}$. Therefore, each row of $P^{(v)}$ sums to one, and $P_{ij}^{(v)}$ can be interpreted as the one-step transition probability from the i-th sample to the j-th anchor in the v-th view. Since $P^{(v)} \in \mathbb{R}^{n \times m}$ is not a square stochastic matrix, we use the anchor-anchor transition probability matrix to characterize high-order sample-anchor relationships. Let $P_{\texttt{aa}}^{(v)} \in \mathbb{R}^{m \times m}$ denote the rows of $P^{(v)}$ corresponding to the selected anchors in the same order as the anchor columns. Using $P_{\texttt{aa}}^{(v)}$, we define the a-th order sample-anchor transition matrix $P_a^{(v)}$ and aggregate the first A orders to obtain $\widetilde{P}^{(v)}$:

$$P_a^{(v)} = P^{(v)} \left(P_{\texttt{aa}}^{(v)}\right)^{a-1}, (a = 1, 2, \ldots, A), \quad \widetilde{P}^{(v)} = \sum_{a=1}^{A} P_a^{(v)} = P^{(v)} \sum_{a=0}^{A-1} \left(P_{\texttt{aa}}^{(v)}\right)^{a}. \tag{3}$$

For $a = 1$, $P_a^{(v)}$ retains the original sample-anchor connections; for $a > 1$, it captures successive transitions among anchors. Finally, we stack the view-specific representations $\widetilde{P}^{(v)}$ of all views into a third-order tensor: $\boldsymbol{\mathcal{P}}_A^{(v)} = \widetilde{P}^{(v)}$, where $\boldsymbol{\mathcal{P}}_A \in \mathbb{R}^{n\times m\times V}$ is named the cross-order anchor transition tensor. Compared with conventional $n \times n \times V$ transition tensors, $\boldsymbol{\mathcal{P}}_A$ preserves multi-order structural information with substantially lower storage and computational costs.

3.2 Non-convex tensor rank approximation

Although TNN is a tractable convex surrogate of tensor rank, its uniform shrinkage may over-penalize dominant singular values and consequently distort the intrinsic low-rank structure (Peng et al., 2026; Liu et al., 2026). To alleviate this bias, we introduce an arctangent-based nonconvex surrogate that strongly suppresses small singular values while applying relatively mild penalties to large ones. The non-convex tensor rank approximation is defined as follows:

$$\|\boldsymbol{\mathcal{Z}}\|_{\text{atan}} = \sum_{v=1}^{V}\sum_{i=1}^{N} \arctan(d\sigma_i(\widehat{\boldsymbol{\mathcal{Z}}}^{(v)})), \tag{4}$$

Specifically, the parameter $d > 0$ controls the strength of the penalty. For a fixed d, the derivative $d/(1 + d^2\sigma_i(\widehat{\boldsymbol{\mathcal{Z}}}^{(v)})^2)$ decreases with $\sigma_i(\widehat{\boldsymbol{\mathcal{Z}}}^{(v)})$, yielding weaker marginal penalization of larger singular values. Increasing d steepens the penalty near zero and accelerates its saturation.

3.3 View-relation graph construction

Different view pairs may exhibit different degrees of structural agreement. We therefore measure view similarity using their transition structures at corresponding orders, so that discrepancies across orders do not cancel during aggregation. Specifically, the distance between the s-th and r-th views is defined as $d_{sr} = \sum_{a=1}^{A} \|\boldsymbol{\mathcal{P}}_a^{(s)} - \boldsymbol{\mathcal{P}}_a^{(r)}\|_F^2$, where $s, r = 1, \ldots, V$. Then the affinity between two different views is defined as $a_{sr} = \exp(-d_{sr}/\sigma^2), (s \neq r, a_{ss} = 1)$, where $\sigma^2 = \eta\,\text{mean}_{s<r}(d_{sr})$, and $\eta > 0$ is a bandwidth scaling factor. Let $\mathbf{A} = [a_{sr}] \in \mathbb{R}^{V\times V}$ be the view affinity matrix and $\mathbf{D}$ be a diagonal matrix with $D_{ss} = \sum_{r=1}^{V} a_{sr}$. Thus, the view graph Laplacian is then defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. The view-relation regularization term can be equivalently written as:

$$\frac{1}{2}\sum_{sr=1}^{V} a_{sr}\|\boldsymbol{\mathcal{Z}}^{(s)} - \boldsymbol{\mathcal{Z}}^{(r)}\|_F^2 = \text{Tr}\left(\ddot{\mathbf{Z}}\mathbf{L}\ddot{\mathbf{Z}}^T\right), \tag{5}$$

where $\ddot{\mathbf{Z}} = [\text{vec}(\boldsymbol{\mathcal{Z}}^{(1)}), \ldots, \text{vec}(\boldsymbol{\mathcal{Z}}^{(V)})] \in \mathbb{R}^{nm\times V}$. This regularizer encourages views with similar transition structures to have consistent recovered representations, while imposing weaker coupling on dissimilar views. It thus complements the global low-rank regularization with pair-specific structural constraints.

3.4 Overall Objective

To mitigate noise and outliers, we decompose $\boldsymbol{\mathcal{P}}_A = \boldsymbol{\mathcal{Z}} + \boldsymbol{\mathcal{E}}$, where the low-rank component $\boldsymbol{\mathcal{Z}}$ capture the shared clustering structure and the sparse component $\boldsymbol{\mathcal{E}}$ absorbs corruptions and view-specific noises. Combining the nonconvex tensor regularizer with view-relation regularization yields the following model:

$$\min_{\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{Z}}\|_{\text{atan}} + \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \frac{\gamma}{2}\sum_{s,r=1}^{V} a_{sr}\|\boldsymbol{\mathcal{Z}}^{(s)} - \boldsymbol{\mathcal{Z}}^{(r)}\|_F^2 \qquad s.t.\ \boldsymbol{\mathcal{P}}_A = \boldsymbol{\mathcal{Z}} + \boldsymbol{\mathcal{E}},\ \boldsymbol{\mathcal{Z}} \geq 0. \tag{6}$$

The nonnegative constraint $\boldsymbol{\mathcal{Z}} \geq 0$ is imposed to preserve the probabilistic interpretation of the learned anchor transition representation. The positive parameters λ and γ control the importance of the sparse error term and the view-relation regularizer, respectively.

4 THEORETICAL ANALYSIS

4.1 ALTERNATING OPTIMIZATION

We employ the alternating direction method of multipliers (ADMM) to optimize Eq. (6) and introduce auxiliary variables $\boldsymbol{\mathcal{J}}$ and $\boldsymbol{\mathcal{G}}$ to obtain the following augmented Lagrangian function:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\mathcal{J}},\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{G}},\boldsymbol{\mathcal{E}},\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{Y}}_2,\boldsymbol{\mathcal{Y}}_3,\rho) &= \|\boldsymbol{\mathcal{J}}\|_{\text{atan}} + \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \frac{\gamma}{2}\sum_{s,r=1}^{V} a_{sr}\|\boldsymbol{\mathcal{Z}}^{(s)} - \boldsymbol{\mathcal{Z}}^{(r)}\|_F^2 \\ &+ \frac{\rho}{2}\|\boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\rho}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{J}} + \boldsymbol{\mathcal{Y}}_2/\rho\|_F^2 + \frac{\rho}{2}\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{G}} + \boldsymbol{\mathcal{Y}}_3/\rho\|_F^2,\ s.t.\ \boldsymbol{\mathcal{G}} \geq 0,\end{aligned} \tag{7}$$

where $\boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{Y}}_2, \boldsymbol{\mathcal{Y}}_3$ are Lagrange multipliers, $\rho > 0$ is the penalty parameter. Then, we alternately optimize each variable while keeping others fixed.

Update $\boldsymbol{\mathcal{J}}$: The subproblem of $\boldsymbol{\mathcal{J}}$ is:

$$\min_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}} + \frac{\rho}{2}\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{J}} + \boldsymbol{\mathcal{Y}}_2/\rho\|_F^2, \tag{8}$$

By defining $\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{Z}} + \frac{\boldsymbol{\mathcal{Y}}_2}{\rho}$ and applying FFT and slice-wise SVD, Theorem 3.2 of (Peng et al., 2015) allows us to reduce Eq. (8) to independent singular-value updates:

$$\sigma_i^{(v)\star} \in \arg\min_{\sigma_i^{(v)} \geq 0} \phi_{\sigma_i^{R(v)}}(\sigma_i^{(v)}) := \frac{1}{\rho}\arctan(d\sigma_i^{(v)}) + \frac{1}{2}(\sigma_i^{(v)} - \sigma_i^{R(v)})^2, \tag{9}$$

where $\sigma_i^{R(v)}$ is the i-th singular value of $\widehat{\boldsymbol{\mathcal{R}}}^{(v)}$, and $\sigma_i^{(v)}$ denotes the corresponding singular value to be updated in $\widehat{\boldsymbol{\mathcal{J}}}^{(v)}$. Any positive minimizer satisfies $\phi'_{\sigma_i^{R(v)}}(\sigma_i^{(v)}) = 0$, yielding

$$(\sigma_i^{(v)})^3 - \sigma_i^{R(v)}(\sigma_i^{(v)})^2 + \frac{1}{d^2}\sigma_i^{(v)} + \frac{1}{d^2}(\frac{d}{\rho} - \sigma_i^{R(v)}) = 0. \tag{10}$$

The following proposition characterizes the solution.

Proposition 4.1. *Under the condition that $d\sigma_i^{R(v)} \leq \sqrt{3}$, the scalar minimization problem has a unique solution. Specifically, $\sigma_i^{(v)\star} = 0$ when $\sigma_i^{R(v)} \leq d/\rho$; otherwise, $\sigma_i^{(v)\star} \in (0, \sigma_i^{R(v)})$ is the unique positive root of the cubic equation in Eq.* (10)*.*

The proof is given in Appendix A. For $\sigma_i^{R(v)} > \frac{\sqrt{3}}{d}$, let Ω_+ denote the nonnegative real roots of the cubic equation in Eq. (10), obtainable using Cardano's formula. The global minimizer is selected by

$$\sigma_i^{(v)\star} \in \arg\min_{\sigma_i^{(v)} \in \{0\}\cup\Omega_+} \phi_{\sigma_i^{R(v)}}(\sigma_i^{(v)}). \tag{11}$$

Finally, we reconstruct each Fourier slice using the updated singular values and the original singular vectors, and obtain $\boldsymbol{\mathcal{J}}$ by $\boldsymbol{\mathcal{J}} = \textbf{ifft}(\widehat{\boldsymbol{\mathcal{J}}})$. The complete update is summarized in Algorithm 1.

Update $\boldsymbol{\mathcal{E}}$: Fixing the other variables, the subproblem associated with $\boldsymbol{\mathcal{E}}$ is:

$$\min_{\boldsymbol{\mathcal{E}}} \lambda\|\boldsymbol{\mathcal{E}}\|_{2,1} + \frac{\rho}{2}\|\boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{Y}}_1/\rho\|_F^2. \tag{12}$$

Let $\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\rho}$, and denote the v-th frontal slices of $\boldsymbol{\mathcal{Q}}$ and $\boldsymbol{\mathcal{E}}$ by $\mathbf{Q}^{(v)} = \boldsymbol{\mathcal{Q}}(:,:,v)$ and $\mathbf{E}^{(v)} = \boldsymbol{\mathcal{E}}(:,:,v)$, respectively. Concatenating these slices along the column dimension gives $\overline{\mathbf{Q}} = \left[\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \ldots, \mathbf{Q}^{(V)}\right] \in \mathbb{R}^{n\times mV}$, with $\overline{\mathbf{E}}$ defined analogously. Thus $\|\boldsymbol{\mathcal{E}}\|_{2,1} = \sum_{i=1}^{n}\|\overline{\mathbf{E}}_{i,:}\|_2$, where$\overline{\mathbf{E}}_{i,:}$ denotes the i-th row of $\overline{\mathbf{E}}$. The problem then admits the following row-wise shrinkage solution (Yuan & Lin, 2005): $\overline{\mathbf{E}}_{i,:} = \left(1 - \frac{1}{\rho\|\overline{\mathbf{Q}}_{i,:}\|_2}\right)_+ \overline{\mathbf{Q}}_{i,:}$.

Here,$(x)_+ = \max(x, 0)$ and $\overline{\mathbf{E}}_{i,:} = 0$ when $\|\overline{\mathbf{Q}}_{i,:}\|_2 = 0$. Finally, $\overline{\mathbf{E}}$ is split into V blocks of m columns and stacked back into $\boldsymbol{\mathcal{E}}$.

Update $\mathcal{Z}$: Fixing the other variables, the subproblem associated with $\mathcal{Z}$ is given by:

$$\min_{\mathcal{Z}} \frac{\gamma}{2} \sum_{s,r=1}^{V} a_{sr} \|\mathcal{Z}^{(s)} - \mathcal{Z}^{(r)}\|_F^2 + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{B}_1\|_F^2 + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{B}_2\|_F^2 + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{B}_3\|_F^2, \tag{13}$$

where $\mathcal{B}_1 = \mathcal{P}_A - \mathcal{E} + \frac{\mathcal{Y}_1}{\rho}, \mathcal{B}_2 = \mathcal{J} - \frac{\mathcal{Y}_2}{\rho}$, and $\mathcal{B}_3 = \mathcal{G} - \frac{\mathcal{Y}_3}{\rho}$. The problem decomposes into mn independent V-dimensional subproblems, whose optimality conditions form linear systems that can be solved simultaneously. Let $\mathcal{B} = \mathcal{B}_1 + \mathcal{B}_2 + \mathcal{B}_3$, and reshape it into $\mathbf{B} \in \mathbb{R}^{nm \times V}$ using the same ordering as Z_v. Setting the gradient with respect to Z_v to zero yields $\mathbf{Z}_v = \rho\mathbf{B}\left(3\rho\mathbf{I}_V + 2\gamma\mathbf{L}\right)^{-1}$, where I_V is the $V \times V$ identity matrix. Finally, Z_v is folded back into $\mathcal{Z}$.

Update $\mathcal{G}$ and multipliers: The $\mathcal{G}$-update is the element-wise nonnegative projection $\mathcal{G} = \max\left(\mathcal{Z} + \frac{\mathcal{Y}_3}{\rho}, 0\right)$. The multipliers are updated as $\mathcal{Y}_1 = \mathcal{Y}_1 + \rho(\mathcal{P}_A - \mathcal{Z} - \mathcal{E})$, $\mathcal{Y}_2 = \mathcal{Y}_2 + \rho(\mathcal{Z} - \mathcal{J})$, $\mathcal{Y}_3 = \mathcal{Y}_3 + \rho(\mathcal{Z} - \mathcal{G})$, followed by $\rho = \min(\kappa\rho, \rho_{\max})$, where $\kappa > 1$. After convergence, we apply k-means to the fused representation $\mathcal{Z}_{\text{fused}} = \sum_{v=1}^{V} \mathcal{Z}^{(v)}$ to obtain the final clustering result.

4.2 Complexity Analysis

It requires $O(nmD + Vnmlog(m) + Vnm^2)$ to construct the cross-order anchor transition tensor, where $D = \sum_{v=1}^{V} d_v$. Following Wu et al. (2019), we rotate $\mathcal{J}$ from $n \times m \times V$ to $n \times V \times m$, the $\mathcal{J}$-update costs $O(Vnm\log(m) + nmV^2)$ for FFT, IFFT and slice-wise SVD, while the $\mathcal{Z}$-update costs $O(V^3 + nmV^2)$. All remaining updates require $O(nmV)$. The overall time complexity of our algorithm remains $O(nmD + Vnmlog(m) + nm^2V + nmV^2)$, which scales linearly with n.

4.3 Convergence Analysis

Convergence of the optimization algorithm is theoretically ensured by Theorems 4.1 and 4.2. The proof is provided in Appendix C and D.

Theorem 4.1. *Sequences $\{\mathcal{Z}_{(t)}\}, \{\mathcal{J}_{(t)}\}, \{\mathcal{G}_{(t)}\}, \{\mathcal{E}_{(t)}\}, \{\mathcal{Y}_{1(t)}\}, \{\mathcal{Y}_{2(t)}\}$, and $\{\mathcal{Y}_{3(t)}\}$ generated by HONR are bounded under the assumptions that $\sum_{\tau=0}^{\infty} \rho_{(\tau)}^{-2}\rho_{(\tau+1)} < \infty$ and $\sum_{\tau=0}^{\infty} = \rho_{(\tau)}^{-1} < \infty$.*

Theorem 4.2. *Let $\{\mathcal{Z}_{(t)}, \mathcal{J}_{(t)}, \mathcal{G}_{(t)}, \mathcal{E}_{(t)}, \mathcal{Y}_{1(t)}, \mathcal{Y}_{2(t)}, \mathcal{Y}_{3(t)}\}$ be a sequence generated by our algorithm. Then, under the assumptions that $\sum_{\tau=0}^{\infty} \rho_{(\tau)}^{-2}\rho_{(\tau+1)} < \infty$ and $\sum_{\tau=0}^{\infty} \rho_{(\tau)}^{-1} < \infty$, the sequence $\{\mathcal{Z}_{(t)}, \mathcal{J}_{(t)}, \mathcal{G}_{(t)}, \mathcal{E}_{(t)}, \mathcal{Y}_{1(t)}, \mathcal{Y}_{2(t)}, \mathcal{Y}_{3(t)}\}$ has at least one accumulation point. For any accumulation point, denoted as $\{\mathcal{Z}_{(*)}, \mathcal{J}_{(*)}, \mathcal{G}_{(*)}, \mathcal{E}_{(*)}, \mathcal{Y}_{1(*)}, \mathcal{Y}_{2(*)}, \mathcal{Y}_{3(*)}\}$, $\{\mathcal{Z}_{(*)}, \mathcal{E}_{(*)}\}$ is a stationary point of optimization problem Eq.* (6).

5 Experiment

Datasets and evaluation matrics: We conduct experiments on seven benchmark datasets, including MSRC (Winn & Jojic, 2005), Flower17 (Li et al., 2023), UCI_digit (Zhang et al., 2023), MITIndoor (Quattoni & Torralba, 2009), Caltech101-all (Zhang et al., 2023), ALOI (Lao et al., 2024) and VGGface50 (Winn & Jojic, 2005). A summary of these datasets is provided in Table 1. To comprehensively assess the clustering performance, seven widely used evaluation metrics are adopted, including clustering accuracy (ACC), normalized mutual information (NMI), Purity, F-score, Precision, Recall and adjusted Rand index (ARI), while higher values indicating better clustering performance.

Dataset	Sample	View	Cluster
MSRC	210	6	7
Flower17	1,360	7	17
UCI-digit	2,000	3	10
MITIndoor	5,360	4	67
Caltech101-all	9,144	5	102
ALOI	10,800	4	100
VGGface50	16,936	4	50

Table 1: Multi-view Datasets

Baselines: Eight baseline methods are selected for comparison, namely LMVSC (Kang et al., 2019), FPMVS-CAG (Wang et al., 2022), FMCNOF (Yang et al., 2021), SFMC (Li et al., 2022), ERMC

(Yang et al., 2022), OMVCDR (Wan et al., 2025), UDBGL (Fang et al., 2024) and DMAC (Wang et al., 2025).

5.1 Experimental Settings

For HONR, we search λ and γ over $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$, and d over $\{0.7, 0.9, 1.1, 1.3, 1.5\}$. The maximum order A and Gaussian bandwidth scaling factor η are set to 3 and 0.3, respectively. To construct the anchor graph, we choose the local neighbor parameter k from $\{2, 3, 4, 5, 6\}$. The number of anchors m is selected from $\{4c, 5c, \ldots, 8c\}$ for the small-scale datasets MSRC, Flower17, UCI_digit and MITIndoor, where c denotes the number of clusters. For the medium-scale datasets Caltech101-all and ALOI, we set $m \in \{4c, 5c\}$, while $m = 5c$ is adopted for the large-scale VGGface50 dataset. Throughout the experiments, ρ_0 and κ are fixed at 0.001 and 1.5, respectively. For a fair comparison, we tune the number of anchors for each anchor-based baseline over the same dataset-specific candidate values used for HONR, while the remaining parameters are set according to the original papers. For all methods, the final clustering step is repeated independently 10 times, and the mean performance and standard deviation are reported in Table 2. **Bold** and underlined values indicated the best and the second-best performance in each case.

5.2 Clustering Results

Table 2 reports the clustering results of HONR and eight competing methods on seven benchmark datasets. Overall, HONR achieves the best performance on 40 out of 49 dataset-metric combinations, demonstrating its effectiveness across datasets with different scales and characteristics. Specifically, HONR ranks first on all seven metrics on MSRC and six metrics on each of Flower17, UCI-digit, MITIndoor, and ALOI. On MSRC, our method achieves an ACC of 89.43% and an NMI of 80.38%, outperforming the corresponding second-best results by 12.29% and 10.57%, respectively. On UCI-digit and MITIndoor, it obtains ACC values of 92.65% and 38.15%, improving upon the second-best result by 8.90 and 11.78 percentage points, respectively. A more substantial improvement can be observed on ALOI, where HONR achieves an ACC of 72.89%, exceeding the strongest baseline by 16.83%. Although HONR does not lead on every metric on Caltech101-all, it achieves the best NMI, Purity, Precision, and ARI. FPMVS-CAG obtains the highest ACC and F-score on this dataset, but its advantage does not generalize consistently across datasets: HONR achieves a higher ACC on all six remaining datasets and a higher F-score on five of them. Moreover, on the large-scale VGGface50 dataset containing 16,936 samples, HONR still achieves the best results in ACC, NMI, Purity, Precision, and ARI. Overall, these results demonstrate the effectiveness of HONR across the evaluated datasets. The contributions of high-order neighbor information and the two regularization terms are further examined in the following ablation studies.

5.3 Ablation Study

(1) Effectiveness of Regularization Terms: We evaluate two variants of HONR: HONR-TNN, which replaces the proposed arctangent-based nonconvex tensor rank approximation with TNN, and HONR-w/o-VR, which removes the view-relation regularization term from Eq. (6). As shown in Fig. 1, the full model generally outperforms both variants on six datasets. Replacing the nonconvex rank approximation with TNN leads to a particularly noticeable decrease in ACC on MITIndoor, while removing the view-relation regularization also reduces performance. These results support the contributions of both regularization terms to clustering performance.

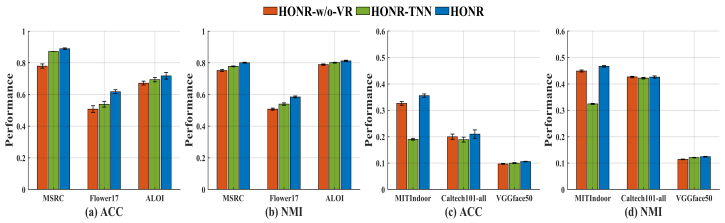


Figure 1: Ablation study of the regularization terms.

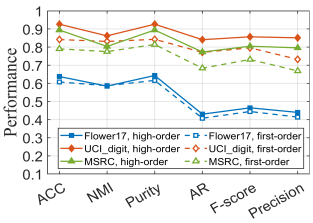


Figure 2: Effect of high-order neighbor information.

Table 2: The clustering performance of HONR and eight baselines on seven datasets.

Dataset	Algorithm	ACC	NMI	Purity	F-score	Precision	Recall	ARI
MSRC	LMVSC	34.24±1.45	26.89±1.97	36.48±1.21	26.91±1.13	21.34±0.88	36.46±2.17	11.41±1.31
	FPMVS-CAG	75.29±6.94	67.27±2.75	76.29±4.82	66.39±4.15	61.49±5.75	72.34±1.54	60.41±5.18
	FMCNOF	48.33±7.50	42.25±7.06	49.67±7.32	39.35±5.99	31.75±6.46	52.74±7.04	26.57±7.68
	SFMC	71.90±0.00	64.85±0.00	72.38±0.00	58.65±0.00	50.31±0.00	70.31±0.00	50.67±0.00
	ERMC	74.29±8.36	69.81±7.19	77.48±6.79	66.43±8.33	63.93±9.34	69.22±7.39	60.73±9.81
	OMVCDR	66.19±0.00	58.23±0.00	67.14±0.00	53.53±0.00	52.46±0.00	54.65±0.00	45.87±0.00
	UDBGL	77.14±0.00	69.03±0.00	77.14±0.00	65.06±0.00	57.96±0.00	74.15±0.00	58.62±0.00
	DMAC	47.90±6.69	42.44±8.20	50.52±7.60	38.73±7.22	34.55±6.49	44.55±9.80	27.45±8.29
	HONR	**89.43±0.30**	**80.38±0.46**	**89.43±0.30**	**80.48±0.58**	**79.59±0.63**	**81.39±0.53**	**77.29±0.68**
Flower17	LMVSC	54.68±0.25	53.43±0.54	55.87±0.73	35.72±0.32	31.68±0.87	40.99±0.76	31.21±0.42
	FPMVS-CAG	32.72±0.00	43.74±0.00	34.49±0.00	28.47±0.00	20.60±0.00	46.09±0.00	22.22±0.00
	FMCNOF	23.44±4.14	29.25±6.22	23.96±4.30	19.67±2.85	12.98±2.38	42.04±5.48	11.83±3.48
	SFMC	27.87±0.00	33.61±0.00	28.53±0.00	18.44±0.00	11.44±0.00	47.48±0.00	10.01±0.00
	ERMC	46.37±1.93	46.29±1.41	48.23±2.03	32.75±1.77	31.67±1.68	33.93±1.94	28.45±1.88
	OMVCDR	57.57±0.00	56.31±0.00	58.60±0.00	43.68±0.00	42.76±0.00	44.64±0.00	40.12±0.00
	UDBGL	35.96±0.00	43.02±0.00	36.91±0.00	17.02±0.00	10.15±0.00	**52.78±0.00**	08.06±0.00
	DMAC	29.10±5.36	29.66±6.25	30.99±5.66	17.88±4.03	17.36±3.93	18.45±4.14	12.65±4.29
	HONR	**63.77±1.21**	**58.47±0.73**	**64.33±1.24**	**46.50±1.18**	**43.98±1.17**	49.34±1.26	**43.00±1.26**
UCI-digit	LMVSC	78.23±4.49	78.36±2.21	81.87±2.78	71.96±3.72	67.74±4.50	76.83±3.38	68.64±4.19
	FPMVS-CAG	67.85±5.13	71.45±3.27	68.62±5.30	63.58±5.02	56.80±7.67	72.89±0.25	58.93±6.00
	FMCNOF	44.95±5.92	50.79±3.81	46.22±5.01	40.54±4.13	30.47±4.66	61.98±4.91	31.34±5.36
	SFMC	76.95±0.00	76.34±0.00	76.95±0.00	66.79±0.00	58.64±0.00	77.58±0.00	62.54±0.00
	ERMC	62.98±6.42	60.48±5.49	65.24±5.97	52.51±6.60	51.01±6.90	54.14±6.34	47.08±7.40
	OMVCDR	43.75±0.00	47.08±0.00	48.70±0.00	39.10±0.00	37.13±0.00	41.30±0.00	31.97±0.00
	UDBGL	83.75±0.00	82.12±0.00	83.75±0.00	78.00±0.00	74.98±0.00	81.27±0.00	75.46±0.00
	DMAC	53.73±15.65	51.89±17.73	56.09±15.08	44.66±15.17	40.88±13.56	49.75±18.39	37.97±16.81
	HONR	**92.65±0.00**	**86.21±0.00**	**92.65±0.00**	**85.69±0.00**	**85.12±0.00**	**86.27±0.00**	**84.09±0.00**
MITindoor	LMVSC	26.37±0.80	38.22±0.53	29.12±0.69	10.11±0.28	07.37±0.39	16.18±0.79	08.25±0.32
	FPMVS-CAG	20.03±0.25	38.43±0.26	20.74±0.31	10.70±0.03	06.64±0.02	27.63±0.01	08.53±0.03
	FMCNOF	06.79±1.01	14.73±1.94	06.89±1.01	04.15±0.49	02.25±0.32	29.32±7.97	01.46±0.57
	SFMC	04.29±0.00	07.36±0.00	04.74±0.00	02.99±0.00	01.52±0.00	**92.96±0.00**	00.09±0.00
	ERMC	12.55±2.35	26.43±4.13	13.80±2.52	06.09±1.63	04.91±1.47	08.08±1.86	04.33±1.70
	OMVCDR	19.74±0.00	**50.05±0.00**	21.08±0.00	17.20±0.00	11.03±0.00	39.04±0.00	15.25±0.00
	UDBGL	07.65±0.00	14.50±0.00	08.54±0.00	03.00±0.00	01.53±0.00	81.08±0.00	00.11±0.00
	DMAC	11.25±1.32	22.12±3.39	08.97±1.11	08.97±0.92	06.28±0.87	15.69±4.23	03.42±1.02
	HONR	**38.15±0.51**	47.56±0.33	**40.12±0.47**	**20.27±0.41**	**18.51±0.46**	22.39±0.38	**18.96±0.42**
Caltech101-all	LMVSC	19.05±0.55	36.68±0.32	35.26±0.29	11.03±0.53	12.31±1.06	10.04±0.63	08.72±0.58
	FPMVS-CAG	**27.53±0.99**	34.45±0.53	34.09±0.71	**20.60±1.44**	14.29±1.51	37.51±2.94	17.23±1.63
	FMCNOF	15.29±1.76	18.25±1.41	18.36±2.11	09.93±1.63	06.09±1.07	30.17±10.42	05.57±1.77
	SFMC	18.94±0.00	25.82±0.00	21.37±0.00	05.53±0.00	02.90±0.00	**60.23±0.00**	00.17±0.00
	ERMC	17.57±2.20	31.39±1.25	29.85±1.05	17.97±3.23	18.63±3.30	17.62±3.79	15.67±3.27
	OMVCDR	16.73±0.00	36.08±0.00	30.64±0.00	13.52±0.00	16.64±0.00	11.38±0.00	11.49±0.00
	UDBGL	18.83±0.01	23.48±0.01	23.16±0.01	05.16±0.01	02.71±0.01	53.28±0.17	-00.22±0.01
	DMAC	12.09±1.90	28.34±3.52	26.53±2.84	09.07±1.91	13.62±3.52	06.85±1.34	07.30±2.06
	HONR	22.46±0.96	**44.37±0.40**	**43.42±0.31**	19.28±1.41	**33.41±2.45**	13.55±1.00	**17.95±1.43**
ALOI	LMVSC	03.80±0.03	10.84±0.11	03.98±0.04	01.13±0.03	01.00±0.01	01.30±0.07	00.01±0.01
	FPMVS-CAG	32.20±1.14	63.45±0.64	32.77±1.08	17.46±0.36	10.18±0.25	61.47±0.34	16.03±0.38
	FMCNOF	10.94±1.51	36.98±4.91	11.09±1.55	06.13±1.50	03.27±0.85	52.91±11.17	04.35±1.56
	SFMC	56.06±0.00	70.91±0.00	56.62±0.00	12.90±0.00	07.27±0.00	57.34±0.00	11.34±0.00
	ERMC	31.14±1.14	51.88±0.82	32.99±1.24	20.74±0.62	16.16±0.77	29.02±0.75	19.72±0.64
	OMVCDR	19.34±0.00	53.02±0.00	21.63±0.00	16.76±0.00	11.12±0.00	34.02±0.00	15.49±0.00
	UDBGL	45.05±0.27	61.71±0.18	46.19±0.29	04.87±0.04	02.52±0.02	**74.47±0.21**	03.01±0.04
	DMAC	41.74±3.13	64.29±3.21	45.87±3.29	27.25±3.30	20.09±2.95	42.68±4.20	26.26±3.37
	HONR	**72.89±1.74**	**81.57±0.47**	**74.64±1.55**	**62.31±1.19**	**58.83±1.56**	66.24±0.80	**61.91±1.21**
VGGface50	LMVSC	10.17±0.30	12.40±0.30	10.97±0.27	04.67±0.11	03.33±0.14	07.90±0.62	01.75±0.16
	FPMVS-CAG	09.19±0.36	12.15±0.43	09.52±0.39	**05.69±0.04**	03.58±0.08	13.90±0.73	02.40±0.09
	FMCNOF	05.94±0.33	05.54±0.59	06.22±0.36	04.11±0.13	02.62±0.14	09.70±1.41	00.81±0.20
	SFMC	03.54±0.00	01.51±0.00	03.68±0.00	04.15±0.00	02.12±0.00	**96.00±0.00**	00.01±0.00
	ERMC	06.06±0.71	06.14±1.50	06.93±0.85	03.49±0.52	03.03±0.51	04.13±0.55	01.07±0.59
	OMVCDR	08.08±0.00	11.66±0.00	09.72±0.00	04.30±0.00	03.24±0.00	06.41±0.00	01.53±0.00
	UDBGL	06.44±0.03	08.71±0.07	07.01±0.04	04.18±0.00	02.18±0.00	49.26±1.51	00.13±0.00
	DMAC	05.41±0.38	04.74±0.63	06.22±0.42	02.63±0.13	02.58±0.12	02.68±0.14	00.48±0.12
	HONR	**10.61±0.18**	**12.66±0.10**	**11.61±0.15**	04.89±0.06	**04.79±0.08**	04.99±0.05	**02.79±0.07**

(2) Effects of High-Order Neighbors: We quantitatively evaluate the effectiveness of high-order neighbor information on MSRC, Flower17, and UCI-digit by comparing HONR using only first-order connections with its high-order counterpart. As shown in Fig. 2, incorporating high-order neighbor information consistently improves clustering performance across all reported metrics on the three datasets. These results suggest that high-order neighbor information captures more comprehensive and long-range relationships beyond direct sample-anchor connections, benefiting multi-view clustering.

5.4 Convergence Study

We evaluate the evaluate the convergence behavior of HONR. Without loss of generality, we select MITIndoor and VGGface50 datasets for illustration. The parameters are set as $\{\lambda, \gamma, m, d, k\} = \{0.01, 0.001, 5c, 1.3, 5\}$. Fig. 3 presents the Frobenius distances between the intermediate and final

iterations, where the subscript $(\cdot)_{(t)}$ denotes the iteration index and $(\cdot)_{(*)}$ denotes the final iteration, together with the values of the objective function. As illustrated in Fig. 3, on both datasets, the variable distances decrease rapidly and approach zero within 100 iterations, while the objective function values stabilize after initial fluctuations. These results indicate stable convergence behavior of the proposed optimization algorithm in practice.

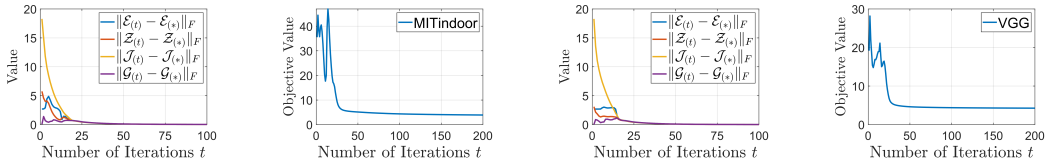


Figure 3: The convergence curves of HONR on MITIndoor (left two panels) and VGGface50 (right two panels).

5.5 Parameter Sensitivity

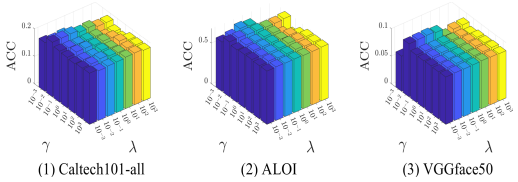


Figure 4: Parameter sensitivity of HONR with respect to λ and γ.

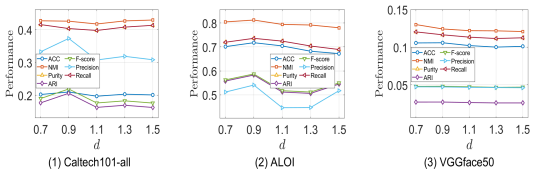


Figure 5: Parameter sensitivity of HONR with respect to d.

(1) Sensitivity to Parameters of λ and γ: In this test, we investigate the sensitivity of HONR to the two trade-off parameters. Without loss of generality, the Caltech101-all, ALOI and VGGface50 datasets are selected for illustration. The ACC values under different parameter combinations is reported in Fig. 4, consistent with the parameter search range adopted in the clustering performance experiments. As can be observed, HONR maintains relatively stable performance over a wide range of parameter combinations, although the best-performing regions differ across datasets. These results indicate that HONR is not overly sensitive to the choices of λ and γ, demonstrating its robustness to parameter variations and facilitating its application in practical scenarios.

(2) Sensitivity to the Arctangent Parameter d: We further test the influence of d in the proposed arctangent-based nonconvex tensor rank approximation on Caltech101-all, ALOI, and VGGface50. The parameter d controls the shape of the arctangent function and thus affects the strength of the non-convex approximation to the tensor rank. We vary d within $\{0.7, 0.9, 1.1, 1.3, 1.5\}$ while fixing the other parameters at their optimal values. As shown in in Fig. 5, the overall clustering performance remains relatively stable within the tested range, which indicate that HONR is generally insensitive to the choice of d, demonstrating the robustness of the proposed arctangent-based rank approximation.

6 Conclusion

We proposed HONR, a scalable multi-view clustering framework for recovering high-order neighbor structure through low-rank tensor learning. By exploiting multi-step random walks, HONR constructs a compact cross-order anchor transition tensor that captures sample-anchor relationships beyond direct connections. An arctangent-based nonconvex tensor rank approximation alleviates excessive shrinkage of dominant singular values, while view-relation regularization strengthens consistency among similar views. We developed an alternating optimization algorithm whose computational complexity scales linearly with the number of samples. The complexity analysis and experimental results on seven benchmark datasets, together with ablation, convergence, and parameter sensitivity studies, demonstrated the effectiveness, robustness, and scalability of the proposed method.

AI USE STATEMENT

Generative AI tools were used to assist with language editing. They were not used for research ideation, method development, code implementation, experimental design or execution, data analysis, or interpretation of results. All AI-assisted content was carefully reviewed and independently verified by the authors, who take full responsibility for the final content of this work.

REFERENCES

Deng Cai and Xinlei Chen. Large scale spectral clustering via landmark-based sparse representation. *IEEE Transactions on Cybernetics*, 45(8):1669–1680, 2015. doi: 10.1109/TCYB.2014.2358564.

Tingquan Deng, Yiying Chen, Taoli Yang, Ge Yang, and Ming Yang. Multi-source partial multi-label learning via tensor decomposition and nonconvex regularization. *Information Fusion*, 112: 102583, 2024.

Guanghao Du, Peng Song, Yuanbo Cheng, Zhaowei Liu, Yanwei Yu, and Wenming Zheng. Weighted tensor-based consistent anchor graph learning for multi-view clustering. *Neurocomputing*, 620:129253, 2025. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2024.129253. URL https://www.sciencedirect.com/science/article/pii/S0925231224020241.

Si-Guo Fang, Dong Huang, Xiao-Sha Cai, Chang-Dong Wang, Chaobo He, and Yong Tang. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):11436–11447, 2024. doi: 10.1109/TNNLS.2023.3261460.

Lele Fu, Zhaoliang Chen, Yongyong Chen, and Shiping Wang. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE Transactions on Multimedia*, 25:4972–4985, 2023. doi: 10.1109/TMM.2022.3185886.

Junwei Han, Kai Xiong, and Feiping Nie. Orthogonal and nonnegative graph reconstruction for large scale clustering. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 1809–1815, 2017. doi: 10.24963/ijcai.2017/251. URL https://doi.org/10.24963/ijcai.2017/251.

Shudong Huang, Ivor W. Tsang, Zenglin Xu, and Jiancheng Lv. Measuring diversity in graph learning: A unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5869–5883, 2022. doi: 10.1109/TKDE.2021.3068461.

Jintian Ji and Songhe Feng. Anchor structure regularization induced multi-view subspace clustering via enhanced tensor rank minimization. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19286–19295, 2023. doi: 10.1109/ICCV51070.2023.01772.

Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. *ArXiv*, abs/1911.09290, 2019. URL https://api.semanticscholar.org/CorpusID:208202248.

Jinghuan Lao, Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. Towards scalable multi-view clustering via joint learning of many bipartite graphs. *IEEE Transactions on Big Data*, 10(1): 77–91, 2024. doi: 10.1109/TBDATA.2023.3325045.

Liang Li, Junpu Zhang, Siwei Wang, Xinwang Liu, Kenli Li, and Keqin Li. Multi-view bipartite graph clustering with coupled noisy feature filter. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12842–12854, 2023. doi: 10.1109/TKDE.2023.3268215.

Pengyuan Li, Dongxia Chang, Zisen Kong, Yiming Wang, and Yao Zhao. Dcmvc: Dual contrastive multi-view clustering. *Neurocomputing*, 635:129889, 2025. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2025.129889. URL https://www.sciencedirect.com/science/article/pii/S0925231225005612.

Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):330–344, 2022. doi: 10.1109/TPAMI.2020.3011148.

Youwei Liang, Dong Huang, Chang-Dong Wang, and Philip S. Yu. Multi-view graph learning by joint modeling of consistency and inconsistency. *IEEE Transactions on Neural Networks and Learning Systems*, 35(2):2848–2862, 2024. doi: 10.1109/TNNLS.2022.3192445.

Chundan Liu, Yongyong Chen, Zhao Kang, Junyu Dong, Panpan Zheng, Chong Peng, and Qiang Cheng. Incomplete multi-view clustering via cross-view diversity enhanced fine-grained tensor completion. *IEEE Trans. Knowl. Data Eng.*, 38(10):6716–6730, 2026. doi: 10.1109/TKDE.2026.3710602. URL https://doi.org/10.1109/TKDE.2026.3710602.

Wei Liu, Junfeng He, and Shih-Fu Chang. Large Graph Construction for Scalable Semi-Supervised Learning. In *International Conference on Machine Learning*, pp. 679–686, 2010. URL https://mlanthology.org/icml/2010/liu2010icml-large/.

Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, and Ce Zhu. S2mvtc: A simple yet efficient scalable multi-view tensor clustering. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24088–24097, 2024. doi: 10.1109/CVPR52733.2024.02274.

Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5249–5257, 2016. doi: 10.1109/CVPR.2016.567.

Feiping Nie, Wei Zhu, and Xuelong Li. Unsupervised large graph embedding. In *AAAI Conference on Artificial Intelligence*, 2017. URL https://api.semanticscholar.org/CorpusID:29163409.

Qiyuan Ou, Siwei Wang, Pei Zhang, Sihang Zhou, and En Zhu. Anchor-based multi-view subspace clustering with hierarchical feature descent. *ArXiv*, abs/2310.07166, 2023. URL https://api.semanticscholar.org/CorpusID:263835310.

Chong Peng, Zhao Kang, Huiqing Li, and Qiang Cheng. Subspace clustering using log-determinant rank approximation. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pp. 925–934, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336642. doi: 10.1145/2783258.2783303. URL https://doi.org/10.1145/2783258.2783303.

Chong Peng, Kehan Kang, Yongyong Chen, Zhao Kang, Chenglizhao Chen, and Qiang Cheng. Fine-grained essential tensor learning for robust multi-view spectral clustering. *IEEE Transactions on Image Processing*, 33:3145–3160, 2024a. doi: 10.1109/TIP.2024.3388969.

Chong Peng, Kai Zhang, Yongyong Chen, Chenglizhao Chen, and Qiang Cheng. Cross-view diversity embedded consensus learning for multi-view clustering. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, IJCAI '24, 2024b. ISBN 978-1-956792-04-1. doi: 10.24963/ijcai.2024/529. URL https://doi.org/10.24963/ijcai.2024/529.

Chong Peng, Chundan Liu, Yang Bai, Yongyong Chen, Zhao Kang, Junyun Dong, Guiyuan Jiang, and Chenglizhao Chen. Fine-grained tensor completion for incomplete multi-view clustering. *Pattern Recognition*, 174:112956, 2026. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2025.112956. URL https://www.sciencedirect.com/science/article/pii/S003132032501619X.

Ariadna Quattoni and Antonio Torralba. Recognizing indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 413–420. IEEE, 2009.

Tomoya Sakai and Atsushi Imiya. Fast spectral clustering with random projection and sampling. In Petra Perner (ed.), *Machine Learning and Data Mining in Pattern Recognition*, pp. 372–384, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-03070-3.

Xinhang Wan, Jiyuan Liu, Xinbiao Gan, Xinwang Liu, Siwei Wang, Yi Wen, Tianjiao Wan, and En Zhu. One-step multi-view clustering with diverse representation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3):5774–5786, 2025. doi: 10.1109/TNNLS.2024.3378194.

Bocheng Wang, Chusheng Zeng, Mulin Chen, and Xuelong Li. Towards learnable anchor for deep multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 21044–21052, 2025.

Shuqin Wang, Yongli Wang, Fang Qiu, Yongyong Chen, Yigang Cen, and Fanghui Zhang. Towards efficient and robust correntropy-based anchor tensor learning for multi-view subspace clustering. *Signal Processing*, 246:110642, 2026. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2026.110642. URL https://www.sciencedirect.com/science/article/pii/S0165168426001568.

Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao, and En Zhu. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31:556–568, 2022. doi: 10.1109/TIP.2021.3131941.

J. Winn and N. Jojic. Locus: learning object classes with unsupervised segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pp. 756–763 Vol. 1, 2005. doi: 10.1109/ICCV.2005.148.

Jianlong Wu, Zhouchen Lin, and Hongbin Zha. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12):5910–5922, 2019. doi: 10.1109/TIP.2019.2916740.

Ben Yang, Xuetao Zhang, Feiping Nie, Fei Wang, Weizhong Yu, and Rong Wang. Fast multi-view clustering via nonnegative and orthogonal factorization. *IEEE Transactions on Image Processing*, 30:2575–2586, 2021. doi: 10.1109/TIP.2020.3045631.

Ben Yang, Xuetao Zhang, Zhiping Lin, Feiping Nie, Badong Chen, and Fei Wang. Efficient and robust multiview clustering with anchor graph regularization. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):6200–6213, 2022. doi: 10.1109/TCSVT.2022.3162575.

Xiaojian You, Haoran Li, Jiali You, and Zhenwen Ren. Consider high-order consistency for multi-view clustering. *Neural Comput. Appl.*, 36(2):717–729, October 2023. ISSN 0941-0643. doi: 10.1007/s00521-023-09054-2. URL https://doi.org/10.1007/s00521-023-09054-2.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1):49–67, 12 2005. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2005.00532.x. URL https://doi.org/10.1111/j.1467-9868.2005.00532.x.

Pei Zhang, Siwei Wang, Liang Li, Changwang Zhang, Xinwang Liu, En Zhu, Zhe Liu, Lu Zhou, and Lei Luo. Let the data choose: flexible and diverse anchor graph fusion for scalable multi-view clustering. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i9.26333. URL https://doi.org/10.1609/aaai.v37i9.26333.

Ruqi Zhang and Zhiwu Lu. Large scale sparse clustering. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pp. 2336–2342. AAAI Press, 2016. ISBN 9781577357704.

Wenhui Zhao, Qin Li, Huafu Xu, Quanxue Gao, Qianqian Wang, and Xinbo Gao. Anchor graph-based feature selection for one-step multi-view clustering. *IEEE Transactions on Multimedia*, 26:7413–7425, 2024. doi: 10.1109/TMM.2024.3367605.

APPENDIX

A PROOF OF PROPOSITION 4.1

Proof. To obtain $\sigma_i^{(v)\star}$, we need to minimize $\phi_{\sigma_i^{R(v)}}(\sigma_i^{(v)})$ under the constraint of $\sigma_i^{(v)} \geq 0$. It is clear that derivative of $\phi_{\sigma_i^{R(v)}}(\sigma_i^{(v)})$ in Eq. (9) and $h(\sigma_i^{(v)}) = (1+d^2\sigma_i^{(v)2})(\sigma_i^{(v)} - \sigma_i^{R(v)}) + \frac{d}{\rho}$ share the same sign. Differentiating h yields

$$h'(\sigma_i^{(v)}) = 3d^2\sigma_i^{(v)2} - 2d^2\sigma_i^{R(v)}\sigma_i^{(v)} + 1. \tag{14}$$

Case 1: If $\sigma_i^{R(v)} = 0$, because $\sigma_i^{(v)} \geq 0$, we always have $\phi'_{\sigma_i^{R(v)}}(\sigma_i^{(v)}) \geq 0$. That is, $\phi_{\sigma_i^{R(v)}}(\sigma_i^{(v)})$ is non-decreasing for any $\sigma_i^{(v)} \geq 0$ and strictly for any $\sigma_i^{(v)} > 0$. Hence the minimizer is $\sigma_i^{(v)\star} = 0$.
Case 2: Suppose $\sigma_i^{R(v)} > 0$ and $\sigma_i^{R(v)} \leq \frac{\sqrt{3}}{d}$. From the discriminant of of $h'(x)$ in Eq. (14), we have $h'(x) \geq 0$ for all $x \geq 0$. Hence, $h(x)$ is strictly increasing on $[0, \infty)$. If $0 < \sigma_i^{R(v)} \leq \frac{d}{\rho}$, we always have $\phi'_{\sigma_i^{R(v)}}(\sigma_i^{(v)}) \geq 0$. The same conclusion holds as in Case 1. If $\sigma_i^{R(v)} > \frac{d}{\rho}$, then there exists at least one root in $(0, \sigma_i^{R(v)})$, because $\phi'_{\sigma_i^{R(v)}}(0) = \frac{d}{\rho} - \sigma_i^{R(v)} < 0$, and $\phi'_{\sigma_i^{R(v)}}(\sigma_i^{R(v)}) > 0$. Since h is non-decreasing, $\phi_{\sigma_i^{R(v)}}(\sigma_i^{(v)}) = 0$ has a unique global minimizer, which is the unique root of $\phi'_{\sigma_i^{R(v)}}(\sigma_i^{(v)}) = 0$ is equivalent to Eq. (10), thus completing the proof. □

B ALGORITHM

Algorithm 1 Update $\mathcal{J}$ via Cardano Formula

Input: Tensor $\mathcal{R}$ and scalar ρ.
Output: $\mathcal{J}$.

```
Ĵ = fft(J, [ ], 3);
for v = 1 : n3 do
    [Û^(v), Ŝ_R̂^(v), V̂^(v)] = SVD(R̂^(v));
    Ŝ_R̂^(v) = diag(σ_1^R(v), σ_2^R(v), ..., σ_N^R(v));
    Obtain σ_i^(v) solving Eq. (10);
    Ŝ_Ĵ^(v) = diag(σ_1^(v), σ_2^(v), ..., σ_N^(v));
    Ĵ^(v) = Û^(v) Ŝ_Ĵ^(v) V̂^(v)T;
end for
J = ifft(Ĵ, [ ], 3)
```

C PROOF OF THEOREM 4.1

Proof. We first establish the boundedness of the multiplier sequence $\{\boldsymbol{\mathcal{Y}}_{1(t)}\}$. By the first-order optimality condition of the $\boldsymbol{\mathcal{E}}$-subproblem at the $(t+1)$-th iteration, we have

$$\lambda \nabla_{\boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{E}}\|_{2,1}|_{\boldsymbol{\mathcal{E}}_{(t+1)}} - \boldsymbol{\mathcal{Y}}_{1(t)} - \rho_{(t)}(\boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{E}}_{(t+1)}) = 0. \tag{15}$$

Together with the multiplier updating $\boldsymbol{\mathcal{Y}}_{1(t+1)} = \boldsymbol{\mathcal{Y}}_{1(t)} + \rho_{(t)}(\boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{E}}_{(t+1)})$, this yields

$$\boldsymbol{\mathcal{Y}}_{1(t+1)} = \lambda \nabla_{\boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{E}}\|_{2,1}|_{\boldsymbol{\mathcal{E}}_{(t+1)}}. \tag{16}$$

Let $E = [E^{(1)}, \ldots, E^{(V)}] \in \mathbb{R}^{n \times mV}$ denote the matrix obtained by concatenating the frontal slices of $\boldsymbol{\mathcal{E}}$, so that $\|\boldsymbol{\mathcal{E}}\|_{2,1} = \sum_{i=1}^{n} \|E_{i,:}\|_2$. For any $D \in \nabla_{\boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{E}}\|_{2,1}$, its i-th row satisfies

$$\begin{cases} D_{i,:} = E_{i,:}/\|E_{i,:}\|_2, & \text{if } E_{i,:} \neq 0, \\ \|D_{i,:}\|_2 \leq 1, & \text{if } E_{i,:} = 0. \end{cases} \tag{17}$$

Consequently, $\|D\|_F^2 = \sum_{i=1}^n \|D_{i,:}\|_2^2 \leq n$, which implies that $\nabla_{\boldsymbol{\mathcal{E}}} \|\boldsymbol{\mathcal{E}}\|_{2,1}|_{\boldsymbol{\mathcal{E}}_{(t+1)}}$ is bounded. Then Eq. (16) implies that $\{\boldsymbol{\mathcal{Y}}_{1(t+1)}\}$ is also bounded.

Next, we establish the boundedness of $\{\boldsymbol{\mathcal{Y}}_{2(t)}\}$. The first-order optimality condition of the $\boldsymbol{\mathcal{J}}$-subproblem gives

$$\nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}}|_{\boldsymbol{\mathcal{J}}_{(t+1)}} + \rho_{(t)} \left(\boldsymbol{\mathcal{J}}_{(t+1)} - \boldsymbol{\mathcal{Z}}_{(t+1)} - \frac{\boldsymbol{\mathcal{Y}}_{2(t)}}{\rho_{(t)}} \right) = \nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}}|_{\boldsymbol{\mathcal{J}}_{(t+1)}} - \boldsymbol{\mathcal{Y}}_{2(t+1)} = 0, \tag{18}$$

where ∇ denotes the limiting subdifferential, and the first equality follows from the multiplier updating $\boldsymbol{\mathcal{Y}}_{2(t+1)} = \boldsymbol{\mathcal{Y}}_{2(t)} + \rho_{(t)}(\boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{J}}_{(t+1)})$.

Hence, it suffices to show that the subgradients of $\|\boldsymbol{\mathcal{J}}_{(t+1)}\|_{\text{atan}}$ are uniformly bounded. Let

$$\widehat{\boldsymbol{\mathcal{J}}}^{(v)} = \mathbb{U}(\widehat{\boldsymbol{\mathcal{J}}}^{(v)}) \cdot \mathbb{D}\{\sigma_1(\widehat{\boldsymbol{\mathcal{J}}}^{(v)}), \cdots, \sigma_c(\widehat{\boldsymbol{\mathcal{J}}}^{(v)})\} \cdot (\mathbb{V}(\widehat{\boldsymbol{\mathcal{J}}}^{(v)}))^T$$

be the SVD of $\widehat{\boldsymbol{\mathcal{J}}}^{(v)}$. Then, according to Deng et al. (2024), we have

$$\begin{aligned} &\nabla_{\widehat{\boldsymbol{\mathcal{J}}}} \|\boldsymbol{\mathcal{J}}_{(t+1)}\|_{\text{atan}} = \nabla_{\widehat{\boldsymbol{\mathcal{J}}}} \sum\nolimits_{v=1}^{V} \sum\nolimits_{i=1}^{n} \arctan(d\sigma_i(\widehat{\boldsymbol{\mathcal{J}}}^{(v)})) \\ =&\mathbb{U}(\widehat{\boldsymbol{\mathcal{J}}}^{(v)})\mathbb{D}\{\frac{d}{1+d^2\sigma_1(\widehat{\boldsymbol{\mathcal{J}}}^{(v)})^2}, \cdots, \frac{d}{1+d^2\sigma_c(\widehat{\boldsymbol{\mathcal{J}}}^{(v)})^2}\}(\mathbb{V}(\widehat{\boldsymbol{\mathcal{J}}}^{(v)}))^T. \end{aligned}$$

Thus, $\|\{\nabla_{\widehat{\boldsymbol{\mathcal{J}}}} \|\boldsymbol{\mathcal{J}}_{(t+1)}\|_{\text{atan}}\}\|_F^2 = \sum_{v=1}^{V} \sum_{i=1}^{n} (\frac{1}{1+d^2\sigma_i(\widehat{\boldsymbol{\mathcal{J}}}^{(v)})^2}) \leq mV$. Denote F_V as the Discrete Fourier transform matrix of size $V \times V$, $\tilde{\mathrm{F}}_V = \frac{1}{\sqrt{V}}\mathrm{F}_V$, $\tilde{\mathrm{F}}_V^H$ as its conjugate transpose, and $\times_3$ as the mode-3 product Deng et al. (2024). Then, $\widehat{\boldsymbol{\mathcal{J}}} = \boldsymbol{\mathcal{J}} \times_3 \tilde{\mathrm{F}}_V$ and by the chain rule, we may obtain that

$$\left\| \{\nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}}\} \right\|_F^2 = \left\| \{\nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}}\} \times_3 \tilde{\mathrm{F}}_V^H \right\|_F^2 = \left\| \{\nabla_{\widehat{\boldsymbol{\mathcal{J}}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}}\} \right\|_F^2 \leq mV. \tag{19}$$

Thus, $\nabla_{\boldsymbol{\mathcal{J}}} \|\boldsymbol{\mathcal{J}}\|_{\text{atan}}|_{\boldsymbol{\mathcal{J}}_{(t+1)}}$ is bounded, which implies that $\{\boldsymbol{\mathcal{Y}}_{2(t+1)}\}$ is bounded.

We next consider the multiplier sequence $\{\boldsymbol{\mathcal{Y}}_{3(t)}\}$. The update of $\boldsymbol{\mathcal{G}}$ is given by

$$\boldsymbol{\mathcal{G}}_{(t+1)} = \max\left\{ \boldsymbol{\mathcal{Z}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{3(t)}}{\rho_{(t)}}, 0 \right\}, \tag{20}$$

where the maximum is taken elementwise.

Define $\mathbb{P}$ and $\bar{\mathbb{P}}$ as two sets that contain index information, such that $\{i,j,v\} \in \mathbb{P}$ if $(\boldsymbol{\mathcal{Z}}_{(t+1)}^{(v)})_{ij} + ((\boldsymbol{\mathcal{Y}}_3)_{(t)}^{(v)})_{ij}/\rho_{(t)} \geq 0$ and $\{i,j,v\} \in \bar{\mathbb{P}}$ if $(\boldsymbol{\mathcal{Z}}_{(t+1)}^{(v)})_{ij} + ((\boldsymbol{\mathcal{Y}}_3)_{(t)}^{(v)})_{ij}/\rho_{(t)} < 0$, respectively. Then, we may divide the $\boldsymbol{\mathcal{G}}$-minimization problem into two parts corresponding to $\mathbb{P}$ and $\bar{\mathbb{P}}$.

On $\mathbb{P}$, we have $\rho_{(t)}(\boldsymbol{\mathcal{G}}_{(t+1)}^{(v)} - \boldsymbol{\mathcal{Z}}_{(t+1)}^{(v)} - (\boldsymbol{\mathcal{Y}}_3)_{(t)}^{(v)}/\rho_{(t)})_{ij} = 0$. By the updating rule of $(\boldsymbol{\mathcal{Y}}_3)_{(t+1)} = (\boldsymbol{\mathcal{Y}}_3)_{(t)} + \rho_{(t)}(\boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{G}}_{(t+1)})$, we have

$$\rho_{(t)}(\boldsymbol{\mathcal{G}}_{(t)}^{(v)} - \boldsymbol{\mathcal{Z}}_{(t+1)}^{(v)} - \frac{(\boldsymbol{\mathcal{Y}}_3)_{(t)}^{(v)}}{\rho_{(t)}})_{ij} = -((\boldsymbol{\mathcal{Y}}_3)_{(t+1)})_{ij}^{(v)} = 0. \tag{21}$$

On $\bar{\mathbb{P}}$, according to the Karush-Kuhn-Tucker (KKT) slackness condition, there exists bounded $\Omega_{ij}^{(v)} \geq 0$, such that

$$\rho_{(t)}(\boldsymbol{\mathcal{G}}_{(t+1)}^{(v)} - \boldsymbol{\mathcal{Z}}_{(t+1)}^{(v)} - (\boldsymbol{\mathcal{Y}}_3)_{(t)}^{(v)}/\rho_{(t)})_{ij} - \Omega_{ij}^{(v)} = -((\boldsymbol{\mathcal{Y}}_3)_{(t+1)}^{(v)})_{ij} - \Omega_{ij}^{(v)} = 0. \tag{22}$$

Thus, by summarizing Eqs. (21) and (22), $((\boldsymbol{\mathcal{Y}}_3)_{(t+1)}^{(v)})_{ij}$ is bounded on both $\mathbb{P}$ and $\bar{\mathbb{P}}$, which suggests that $\{(\boldsymbol{\mathcal{Y}}_3)_{(t+1)}\}$ is bounded.

For the augmented Larange function, we have the following chain of equalities:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t+1)}, \boldsymbol{\mathcal{Y}}_{2(t+1)}, \boldsymbol{\mathcal{Y}}_{3(t+1)}, \rho_{(t+1)}) \\
=&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t)}, \boldsymbol{\mathcal{Y}}_{2(t)}, \boldsymbol{\mathcal{Y}}_{3(t)}, \rho_{(t)}) \\
&+ \frac{\rho_{(t+1)}}{2}\left\|\boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{E}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{1(t+1)}}{\rho_{(t+1)}}\right\|_F^2 + \frac{\rho_{(t+1)}}{2}\left\|\boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{J}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{2(t+1)}}{\rho_{(t+1)}}\right\|_F^2 \\
&+ \frac{\rho_{(t+1)}}{2}\left\|\boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{G}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{3(t+1)}}{\rho_{(t+1)}}\right\|_F^2 - \frac{\rho_{(t)}}{2}\left\|\boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{E}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{1(t)}}{\rho_{(t)}}\right\|_F^2 \\
&- \frac{\rho_{(t)}}{2}\left\|\boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{J}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{2(t)}}{\rho_{(t)}}\right\|_F^2 - \frac{\rho_{(t)}}{2}\left\|\boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{G}}_{(t+1)} + \frac{\boldsymbol{\mathcal{Y}}_{3(t)}}{\rho_{(t)}}\right\|_F^2,
\end{aligned} \tag{23}$$

Define $\boldsymbol{\mathcal{R}}_1 = \boldsymbol{\mathcal{P}}_A - \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{E}}_{(t+1)}$, $R_2 = \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{J}}_{(t+1)}$ and $R_3 = \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{G}}_{(t+1)}$. Then Eq. (23) can be equivalently rewritten as:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t+1)}, \boldsymbol{\mathcal{Y}}_{2(t+1)}, \boldsymbol{\mathcal{Y}}_{3(t+1)}, \rho_{(t+1)}) \\
=&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t)}, \boldsymbol{\mathcal{Y}}_{2(t)}, \boldsymbol{\mathcal{Y}}_{3(t)}, \rho_{(t)}) \\
&+ \frac{\rho_{(t+1)} - \rho_{(t)}}{2}\|\boldsymbol{\mathcal{R}}_1\|_F^2 + \left\langle \boldsymbol{\mathcal{R}}_1,\ \boldsymbol{\mathcal{Y}}_{1(t+1)} - \boldsymbol{\mathcal{Y}}_{1(t)} \right\rangle + \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{1(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{1(t)}\|_F^2 \\
&+ \frac{\rho_{(t+1)} - \rho_{(t)}}{2}\|\boldsymbol{\mathcal{R}}_2\|_F^2 + \left\langle \boldsymbol{\mathcal{R}}_2,\ \boldsymbol{\mathcal{Y}}_{2(t+1)} - \boldsymbol{\mathcal{Y}}_{2(t)} \right\rangle + \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{2(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{2(t)}\|_F^2 \\
&+ \frac{\rho_{(t+1)} - \rho_{(t)}}{2}\|\boldsymbol{\mathcal{R}}_3\|_F^2 + \left\langle \boldsymbol{\mathcal{R}}_3,\ \boldsymbol{\mathcal{Y}}_{3(t+1)} - \boldsymbol{\mathcal{Y}}_{3(t)} \right\rangle + \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{3(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{3(t)}\|_F^2 \\
=&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t)}, \boldsymbol{\mathcal{Y}}_{2(t)}, \boldsymbol{\mathcal{Y}}_{3(t)}, \rho_{(t)}) \\
&+ \frac{\rho_{(t+1)} + \rho_{(t)}}{2\rho_{(t)}^2}\|(\boldsymbol{\mathcal{Y}}_{1(t+1)} - \boldsymbol{\mathcal{Y}}_{1(t)})\|_F^2 + \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{1(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{1(t)}\|_F^2 \\
&+ \frac{\rho_{(t+1)} + \rho_{(t)}}{2\rho_{(t)}^2}\|(\boldsymbol{\mathcal{Y}}_{2(t+1)} - \boldsymbol{\mathcal{Y}}_{2(t)})\|_F^2 + \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{2(t+1)}\|_F^2 - \frac{1}{2\rho_{2(t)}}\|\boldsymbol{\mathcal{Y}}_{2(t)}\|_F^2 \\
&+ \frac{\rho_{(t+1)} + \rho_{(t)}}{2\rho_{(t)}^2}\|(\boldsymbol{\mathcal{Y}}_{3(t+1)} - \boldsymbol{\mathcal{Y}}_{3(t)})\|_F^2 + \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{3(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{3(t)}\|_F^2
\end{aligned} \tag{24}$$

Moreover, since $\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}$ and $\boldsymbol{\mathcal{E}}_{(t+1)}$ are the optimal solutions to their corresponding subproblems, we have

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t+1)}, \boldsymbol{\mathcal{Y}}_{2(t+1)}, \boldsymbol{\mathcal{Y}}_{3(t+1)}, \rho_{(t+1)}) \\
\leq&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t)}, \boldsymbol{\mathcal{J}}_{(t)}, \boldsymbol{\mathcal{G}}_{(t)}, \boldsymbol{\mathcal{E}}_{(t)}, \boldsymbol{\mathcal{Y}}_{1(t)}, \boldsymbol{\mathcal{Y}}_{2(t)}, \boldsymbol{\mathcal{Y}}_{3(t)}, \rho_{(t+1)}) + \frac{\rho_{(t+1)} + \rho_{(t)}}{2\rho_{(t)}^2}\|\boldsymbol{\mathcal{Y}}_{1(t+1)} - \boldsymbol{\mathcal{Y}}_{1(t)}\|_F^2 \\
&+ \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{1(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{1(t)}\|_F^2 + \frac{\rho_{(t+1)} + \rho_{(t)}}{2\rho_{(t)}^2}\|\boldsymbol{\mathcal{Y}}_{2(t+1)} - \boldsymbol{\mathcal{Y}}_{2(t)}\|_F^2 \\
&+ \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{2(t+1)}\|_F^2 - \frac{1}{2\rho_{2(t)}}\|\boldsymbol{\mathcal{Y}}_{2(t)}\|_F^2 + \frac{\rho_{(t+1)} + \rho_{(t)}}{2\rho_{(t)}^2}\|\boldsymbol{\mathcal{Y}}_{3(t+1)} - \boldsymbol{\mathcal{Y}}_{3(t)}\|_F^2 \\
&+ \frac{1}{2\rho_{(t+1)}}\|\boldsymbol{\mathcal{Y}}_{3(t+1)}\|_F^2 - \frac{1}{2\rho_{(t)}}\|\boldsymbol{\mathcal{Y}}_{3(t)}\|_F^2
\end{aligned} \tag{25}$$

Thus, we iterate the above inequality $t+1$ times and obtain the following chain of inequalities:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)}, \boldsymbol{\mathcal{J}}_{(t+1)}, \boldsymbol{\mathcal{G}}_{(t+1)}, \boldsymbol{\mathcal{E}}_{(t+1)}, \boldsymbol{\mathcal{Y}}_{1(t+1)}, \boldsymbol{\mathcal{Y}}_{2(t+1)}, \boldsymbol{\mathcal{Y}}_{3(t+1)}, \rho_{(t+1)}) \\
\leq&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(0)}, \boldsymbol{\mathcal{J}}_{(0)}, \boldsymbol{\mathcal{G}}_{(0)}, \boldsymbol{\mathcal{E}}_{(0)}, \boldsymbol{\mathcal{Y}}_{1(0)}, \boldsymbol{\mathcal{Y}}_{2(0)}, \boldsymbol{\mathcal{Y}}_{3(0)}, \rho_{(0)}) \\
&+ \sum_{\tau=0}^{t}\{\frac{\rho_{(\tau+1)} + \rho_{(\tau)}}{2\rho_{(\tau)}^2}\{\|\boldsymbol{\mathcal{Y}}_{1(\tau+1)} - \boldsymbol{\mathcal{Y}}_{1(\tau)}\|_F^2 + \|\boldsymbol{\mathcal{Y}}_{2(\tau+1)} - \boldsymbol{\mathcal{Y}}_{2(\tau)}\|_F^2 + \|\boldsymbol{\mathcal{Y}}_{3(\tau+1)} - \boldsymbol{\mathcal{Y}}_{3(\tau)}\|_F^2\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2\rho_{(\tau+1)}}\{\|\boldsymbol{\mathcal{Y}}_{1(\tau+1)}\|_F^2+\|\boldsymbol{\mathcal{Y}}_{2(\tau+1)}\|_F^2+\|\boldsymbol{\mathcal{Y}}_{3(\tau+1)}\|_F^2\}\}\\
\leq&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(0)},\boldsymbol{\mathcal{J}}_{(0)},\boldsymbol{\mathcal{G}}_{(0)},\boldsymbol{\mathcal{E}}_{(0)},\boldsymbol{\mathcal{Y}}_{1(0)},\boldsymbol{\mathcal{Y}}_{2(0)},\boldsymbol{\mathcal{Y}}_{3(0)},\rho_{(0)})+6\Gamma\sum_{\tau=0}^{t}\frac{\rho_{(\tau+1)}}{\rho_{(\tau)}^2}+8\Gamma\sum_{\tau=0}^{t+1}\frac{1}{\rho_{(\tau)}},
\end{aligned}
\tag{26}
$$

where Γ is an upper-bound of $\{\|\boldsymbol{\mathcal{Y}}_{1(\tau)}\|_F^2\}, \{\|\boldsymbol{\mathcal{Y}}_{2(\tau)}\|_F^2\}$ and $\{\|\boldsymbol{\mathcal{Y}}_{3(\tau)}\|_F^2\}$. Given bounded initialization of $\boldsymbol{\mathcal{Z}}_{(0)},\boldsymbol{\mathcal{J}}_{(0)},\boldsymbol{\mathcal{G}}_{(0)},\boldsymbol{\mathcal{E}}_{(0)},\boldsymbol{\mathcal{Y}}_{1(0)},\boldsymbol{\mathcal{Y}}_{2(0)},\boldsymbol{\mathcal{Y}}_{3(0)}$ and $\rho_{(0)}$. and according to the assumptions that $\sum_{\tau=0}^{\infty}\rho_\tau^{-1}<\infty$ and $\sum_{\tau=0}^{\infty}\rho_\tau^{-2}\rho_\tau<\infty$, it is clear that

$$
\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)},\boldsymbol{\mathcal{J}}_{(t+1)},\boldsymbol{\mathcal{G}}_{(t+1)},\boldsymbol{\mathcal{E}}_{(t+1)},\boldsymbol{\mathcal{Y}}_{1(t+1)},\boldsymbol{\mathcal{Y}}_{2(t+1)},\boldsymbol{\mathcal{Y}}_{3(t+1)},\rho_{(t+1)})<\infty. \tag{27}
$$

Thus, $\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)},\boldsymbol{\mathcal{J}}_{(t+1)},\boldsymbol{\mathcal{G}}_{(t+1)},\boldsymbol{\mathcal{E}}_{(t+1)},\boldsymbol{\mathcal{Y}}_{1(t+1)},\boldsymbol{\mathcal{Y}}_{2(t+1)},\boldsymbol{\mathcal{Y}}_{3(t+1)},\rho_{(t+1)})$ is bounded, which implies that all terms on the right side of the following equation are bounded:

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{\mathcal{Z}}_{(t+1)},\boldsymbol{\mathcal{J}}_{(t+1)},\boldsymbol{\mathcal{G}}_{(t+1)},\boldsymbol{\mathcal{E}}_{(t+1)},\boldsymbol{\mathcal{Y}}_{1(t+1)},\boldsymbol{\mathcal{Y}}_{2(t+1)},\boldsymbol{\mathcal{Y}}_{3(t+1)},\rho_{(t+1)})\\
&=\|\boldsymbol{\mathcal{J}}_{(t+1)}\|_{\text{atan}}+\lambda\|\boldsymbol{\mathcal{E}}_{(t+1)}\|_{2,1}+\frac{\gamma}{2}\sum_{s,r=1}^{V}a_{sr}\|\boldsymbol{\mathcal{Z}}_{(t+1)}^{(s)}-\boldsymbol{\mathcal{Z}}_{(t+1)}^{(r)}\|_F^2\\
&+\frac{\rho_{(t+1)}}{2}\left\|\boldsymbol{\mathcal{P}}_A-\boldsymbol{\mathcal{Z}}_{(t+1)}-\boldsymbol{\mathcal{E}}_{(t+1)}+\frac{\boldsymbol{\mathcal{Y}}_{1(t+1)}}{\rho_{(t+1)}}\right\|_F^2+\frac{\rho_{(t+1)}}{2}\left\|\boldsymbol{\mathcal{Z}}_{(t+1)}-\boldsymbol{\mathcal{J}}_{(t+1)}+\frac{\boldsymbol{\mathcal{Y}}_{2(t+1)}}{\rho_{(t+1)}}\right\|_F^2\\
&+\frac{\rho_{(t+1)}}{2}\left\|\boldsymbol{\mathcal{Z}}_{(t+1)}-\boldsymbol{\mathcal{G}}_{(t+1)}+\frac{\boldsymbol{\mathcal{Y}}_{3(t+1)}}{\rho_{(t+1)}}\right\|_F^2.
\end{aligned}
\tag{28}
$$

Therefore $\|\boldsymbol{\mathcal{E}}_{(t+1)}\|_{2,1}$ is bounded, then the boundedness of $\boldsymbol{\mathcal{E}}_{(t+1)}$ is straightforward. Moreover, since $\rho_{(t+1)}\geq\rho_0>0$, the boundedness of the first quadratic term implies that $\|\boldsymbol{\mathcal{P}}_A-\boldsymbol{\mathcal{Z}}_{(t+1)}-\boldsymbol{\mathcal{E}}_{(t+1)}+\boldsymbol{\mathcal{Y}}_{1(t+1)}/\rho_{(t+1)}\|_F^2$ is bounded. As $\boldsymbol{\mathcal{P}}_A$ is fixed and both $\boldsymbol{\mathcal{E}}_{(t+1)}$ and $\boldsymbol{\mathcal{Y}}_{1(t+1)}$ are bounded, $\boldsymbol{\mathcal{Z}}_{(t+1)}$ is bounded. Similarly, the boundedness of $\|\boldsymbol{\mathcal{Z}}_{(t+1)}-\boldsymbol{\mathcal{J}}_{(t+1)}+\boldsymbol{\mathcal{Y}}_{2(t+1)}/\rho_{(t+1)}\|_F^2$, together with that of $\boldsymbol{\mathcal{Z}}_{(t+1)}$ and $\boldsymbol{\mathcal{Y}}_{2(t+1)}$, then implies that $\boldsymbol{\mathcal{J}}_{(t+1)}$ is bounded. Finally, the boundedness of the last quadratic term $\|\boldsymbol{\mathcal{Z}}_{(t+1)}-\boldsymbol{\mathcal{G}}_{(t+1)}+\boldsymbol{\mathcal{Y}}_{3(t+1)}/\rho_{(t+1)}\|_F^2$, $\boldsymbol{\mathcal{Z}}_{(t+1)}$ and $\boldsymbol{\mathcal{Y}}_{3(t+1)}$ imply that $\boldsymbol{\mathcal{G}}_{(t+1)}$ is bounded.

Therefore, all variable sequences $\{\boldsymbol{\mathcal{E}}_{(t)}\}$, $\{\boldsymbol{\mathcal{Z}}_{(t)}\}$, $\{\boldsymbol{\mathcal{J}}_{(t)}\}$, and $\{\boldsymbol{\mathcal{G}}_{(t)}\}$ are bounded. Together with the boundedness of the three multiplier sequences, this completes the proof.

□

D PROOF OF THEOREM 4.2

Proof. Under the assumptions in the theorem, and according to Theorem 4.1, it is clear that the sequence $\{\boldsymbol{\mathcal{Z}}_{(t)},\boldsymbol{\mathcal{J}}_{(t)},\boldsymbol{\mathcal{G}}_{(t)},\boldsymbol{\mathcal{E}}_{(t)},\boldsymbol{\mathcal{Y}}_{1(t)},\boldsymbol{\mathcal{Y}}_{2(t)},\boldsymbol{\mathcal{Y}}_{3(t)}\}$ is bounded. Then according to the Bolzano-Weierstrass theorem, the sequence $\{\boldsymbol{\mathcal{Z}}_{(t)},\boldsymbol{\mathcal{J}}_{(t)},\boldsymbol{\mathcal{G}}_{(t)},\boldsymbol{\mathcal{E}}_{(t)},\boldsymbol{\mathcal{Y}}_{1(t)},\boldsymbol{\mathcal{Y}}_{2(t)},\boldsymbol{\mathcal{Y}}_{3(t)}\}$ has at least one accumulation point, which is denoted as $\{\boldsymbol{\mathcal{Z}}_{(*)},\boldsymbol{\mathcal{J}}_{(*)},\boldsymbol{\mathcal{G}}_{(*)},\boldsymbol{\mathcal{E}}_{(*)},\boldsymbol{\mathcal{Y}}_{1(*)},\boldsymbol{\mathcal{Y}}_{2(*)},\boldsymbol{\mathcal{Y}}_{3(*)}\}$. Without loss of generality, we assume that the sequence $\{\boldsymbol{\mathcal{Z}}_{(t)},\boldsymbol{\mathcal{J}}_{(t)},\boldsymbol{\mathcal{G}}_{(t)},\boldsymbol{\mathcal{E}}_{(t)},\boldsymbol{\mathcal{Y}}_{1(t)},\boldsymbol{\mathcal{Y}}_{2(t)},\boldsymbol{\mathcal{Y}}_{3(t)}\}$ itself converges to $\{\boldsymbol{\mathcal{Z}}_{(*)},\boldsymbol{\mathcal{J}}_{(*)},\boldsymbol{\mathcal{G}}_{(*)},\boldsymbol{\mathcal{E}}_{(*)},\boldsymbol{\mathcal{Y}}_{1(*)},\boldsymbol{\mathcal{Y}}_{2(*)},\boldsymbol{\mathcal{Y}}_{3(*)}\}$. Next, we prove that $\{\boldsymbol{\mathcal{Z}}_{(*)},\boldsymbol{\mathcal{E}}_{(*)}\}$ is a stationary point of Eq. (6).

According to the updating rules of $\boldsymbol{\mathcal{Y}}_1,\boldsymbol{\mathcal{Y}}_2$ and $\boldsymbol{\mathcal{Y}}_3$, we have

$$
\begin{aligned}
&\lim_{k\to\infty}\boldsymbol{\mathcal{P}}_A-\boldsymbol{\mathcal{Z}}_{(t)}-\boldsymbol{\mathcal{E}}_{(t)}=\lim_{k\to\infty}((\boldsymbol{\mathcal{Y}}_1)_{(t+1)}-(\boldsymbol{\mathcal{Y}}_1)_{(t)})/\rho_{(t)}=\boldsymbol{\mathcal{P}}_A-\boldsymbol{\mathcal{Z}}_{(*)}-\boldsymbol{\mathcal{E}}_{(*)}=0,\\
&\lim_{k\to\infty}\boldsymbol{\mathcal{Z}}_{(t)}-\boldsymbol{\mathcal{J}}_{(t)}=\lim_{k\to\infty}((\boldsymbol{\mathcal{Y}}_2)_{(t+1)}-(\boldsymbol{\mathcal{Y}}_2)_{(t)})/\rho_{(t)}=\boldsymbol{\mathcal{Z}}_{(*)}-\boldsymbol{\mathcal{J}}_{(*)}=0,\\
&\lim_{k\to\infty}\boldsymbol{\mathcal{Z}}_{(t)}-\boldsymbol{\mathcal{G}}_{(t)}=\lim_{k\to\infty}((\boldsymbol{\mathcal{Y}}_3)_{(t+1)}-(\boldsymbol{\mathcal{Y}}_3)_{(t)})/\rho_{(t)}=\boldsymbol{\mathcal{Z}}_{(*)}-\boldsymbol{\mathcal{G}}_{(*)}=0.
\end{aligned}
$$

By the first-order optimality condition of $\boldsymbol{\mathcal{J}}$, we have

$$
\begin{aligned}
&\lim_{k\to\infty}\nabla_{\boldsymbol{\mathcal{J}}}\|\boldsymbol{\mathcal{J}}\|_{\text{atan}}|_{\boldsymbol{\mathcal{J}}_{(t+1)}}+\rho_{(t)}(\boldsymbol{\mathcal{J}}_{(t+1)}-\boldsymbol{\mathcal{Z}}_{(t+1)}-(\boldsymbol{\mathcal{Y}}_2)_{(t)}/\rho_{(t)})\\
&=\lim_{k\to\infty}\nabla_{\boldsymbol{\mathcal{J}}}\|\boldsymbol{\mathcal{J}}\|_{\text{atan}}|_{\boldsymbol{\mathcal{J}}_{(t+1)}}-(\boldsymbol{\mathcal{Y}}_2)_{(t+1)}=\nabla_{\boldsymbol{\mathcal{J}}}\|\boldsymbol{\mathcal{J}}\|_{\text{atan}}|_{\boldsymbol{\mathcal{J}}_{(*)}}-(\boldsymbol{\mathcal{Y}}_2)_{(*)}=0.
\end{aligned}
$$

By the first-order optimality condition of $\boldsymbol{\mathcal{E}}$, we have

$$\begin{aligned}&\lim_{k\to\infty} \nabla_{\boldsymbol{\mathcal{E}}}\|\boldsymbol{\mathcal{E}}\|_{2,1}|_{\boldsymbol{\mathcal{E}}_{(t+1)}} + \rho_{(t)}(\boldsymbol{\mathcal{E}}_{(t+1)} + \boldsymbol{\mathcal{Z}}_{(t+1)} - \boldsymbol{\mathcal{P}}_A - (\boldsymbol{\mathcal{Y}}_1)_{(t)}/\rho_{(t)})\\ &= \lim_{k\to\infty} \nabla_{\boldsymbol{\mathcal{E}}}\|\boldsymbol{\mathcal{E}}\|_{2,1}|_{\boldsymbol{\mathcal{E}}_{(t+1)}} - (\boldsymbol{\mathcal{Y}}_1)_{(t+1)} = \nabla_{\boldsymbol{\mathcal{E}}}\|\boldsymbol{\mathcal{E}}\|_{2,1}|_{\boldsymbol{\mathcal{E}}_{(*)}} - (\boldsymbol{\mathcal{Y}}_1)_{(*)} = 0.\end{aligned}$$

The other conditions are easy to verify by standard optimization. Therefore, $\{\boldsymbol{\mathcal{Z}}_{(*)}, \boldsymbol{\mathcal{J}}_{(*)}, \boldsymbol{\mathcal{G}}_{(*)}, \boldsymbol{\mathcal{E}}_{(*)}, \boldsymbol{\mathcal{Y}}_{1(*)}, \boldsymbol{\mathcal{Y}}_{2(*)}, \boldsymbol{\mathcal{Y}}_{3(*)}\}$ satisfies the KKT conditions of the augmented Lagrange function in Eq. (7). Thus, $\{\boldsymbol{\mathcal{Z}}_{(*)}, \boldsymbol{\mathcal{E}}_{(*)}\}$ is a stationary point of Eq. (6).

□