Releasing Anchors from Cross-View Correspondence: Probabilistic Multi-View Anchor Graph Clustering

Zhoumin Lu[1] **Yongbo Yu**[2] **Yu Duan**[3] **Feiping Nie**[2] **Yicong Zhou**[1,*]

[1]Department of Artificial Intelligence, University of Macau
[2]School of Artificial Intelligence, OPtics and ElectroNics, Northwestern Polytechnical University
[3]School of Telecommunications Engineering, Xidian University
walker.zhoumin.lu@gmail.com, yongboyu@mail.nwpu.edu.cn, duanyuee@gmail.com,
feipingnie@gmail.com, yicongzhou@um.edu.mo

Abstract

Anchor graphs are a standard route to scalable multi-view clustering, but most existing pipelines still depend on either a shared anchor set sampled from concatenated features or a separate anchor-alignment stage before fusion. Both choices are restrictive: the former hard-codes equal anchor budgets and importance at anchor-construction time, while the latter introduces a cumbersome correspondence problem and still struggles with unequal anchor numbers. We develop an alignment-free probabilistic formulation in which view-specific anchor graphs are noisy manifestations of a shared latent cluster identity rather than objects that require cross-view correspondence. Each view has its own anchor space, number of anchors and cluster prototypes; the only shared latent variable is the sample label. We instantiate this idea with a Gaussian latent-label model and a deterministic prototype M-step. The resulting variational EM algorithm admits closed-form updates with convergence guarantees, while experiments demonstrate competitive clustering performance and practical runtime on most datasets.

1 Introduction

Anchor graphs [Yang and Tan, 2024] are a valuable technique for scaling clustering and graph learning to large datasets because they replace dense sample–sample affinities with sample–anchor relations. In the multi-view setting, however, most anchor-based pipelines still inherit a hidden coordination problem: what should be shared across views? A common engineering shortcut is to concatenate all views and sample one common anchor set. This is convenient, but it commits the method to a single anchor coordinate system, a single anchor budget, and an implicit equal weighting of views during anchor discovery. A more flexible alternative is to construct anchors independently and then align anchor sets before fusion. Yet explicit alignment is itself awkward: it is combinatorial in principle, typically relaxed in practice, and usually presumes equal anchor cardinality. Another line of work performs joint learning across views, which requires all views to share an anchor graph or enhance the consistency among anchor graphs through certain regularizations, thereby indirectly/implicitly achieving anchor alignment. However, graph sharing easily overlooks view-specific complementary information, whereas consistency enhancement often falls back into the previous alignment problem, which can be understood as real-time alignment. Furthermore, this practice is generally more complex and inefficient than the previous two strategies. These issues are particularly visible in existing anchor-based multi-view methods that either emphasize anchor correspondence or improve anchor quality only after some notion of inter-view alignment has already been imposed.

*Corresponding author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

We start from a simpler modeling stance. Anchors are view-local coordinates used to summarize each view efficiently; they need not be semantically shared objects. What should be shared across views is not anchor identity, but the latent cluster label of each sample. Once this viewpoint is adopted, each view may maintain its own anchor system and even its own anchor number m_v, while clustering is performed by coupling views only through latent labels.

This paper develops a probabilistic model embodying that idea. For each view v, the sample-specific anchor vector $\boldsymbol{w}_i^{(v)} \in \mathbb{R}^{m_v}$ is generated from a view-specific Gaussian prototype indexed by a shared latent label z_i. The prototypes are free parameters in each view-specific anchor space rather than low-rank factorizations that would reintroduce cross-view coordination. This choice is particularly natural when anchor vectors are normalized or approximately normalized: then a prototype is simply the weighted average anchor profile of a cluster in that view. Our main contributions are as follows.

- We formulate alignment-free multi-view anchor clustering as latent-label inference with view-specific Gaussian prototypes, thereby eliminating the need for anchor correspondence and naturally permitting unequal anchor numbers.
- We derive a fully closed-form variational inference procedure for all model variables, making the method easy to implement and lightweight to run; its monotonicity, boundedness, and convergence properties are theoretically justified.
- Experiments demonstrate that our model is competitive in both effectiveness and efficiency, benefits from unequal view-specific anchor numbers, and provides a practical way to aggregate independently constructed view-specific anchor graphs.

2 Related Work

Anchor-based scalable multi-view clustering. From the perspective of anchor coordination, existing scalable multi-view anchor clustering methods can be organized into three broad families: (1) Selecting a shared anchor set from concatenated multi-view features and then performing subsequent learning [Xia et al., 2022] or fusion [Li et al., 2022, Lu et al., 2025a]; (2) Learning anchor graphs independently for each view and then either directly concatenating them [Kang et al., 2020, Zhang et al., 2023] or aligning them before fusion [Wang et al., 2022a, Liu et al., 2024]; (3) Jointly learning all view-specific anchor graphs while enhancing cross-view consistency through tensor norms [Ji and Feng, 2025, Kong et al., 2026, Wang et al., 2026], consensus embeddings [Wang et al., 2024], diffusion mechanisms [Zou et al., 2025], structural regularization [Lu et al., 2025b], or by directly sharing a common anchor graph [Wang et al., 2022b, Zhang et al., 2024, Chen et al., 2025, Tang et al., 2026]. These methods are important reference points, but they also make the present gap clear. Most still require either an implicitly shared anchor system, a consensus anchor space, or an explicit alignment/fusion stage. Our formulation is orthogonal to this design choice: anchors are allowed to remain view-local, and cross-view interaction occurs only through the shared latent cluster label.

Variational latent-variable perspective. The proposed inference procedure sits in the classical variational EM and coordinate-ascent variational inference (CAVI) tradition [Dempster et al., 1977, Jordan et al., 1999, Bishop, 2006, Blei et al., 2017]. What is specific here is not generic variational machinery, but its use to produce a fully closed-form, unequal-anchor, alignment-free multi-view anchor clustering model with deterministic prototype updates and complete convergence analysis. To our knowledge, that combination is not spelled out in the current anchor-clustering literature.

3 Generative Model with Free View-Specific Prototypes

3.1 Problem setting

We observe n samples from V views. View v provides an observable sample-to-anchor affinity/assignment vector $\boldsymbol{w}_i^{(v)} \in \mathbb{R}^{m_v}$, for $i = 1, \ldots, n$, where m_v is the number of anchors used in view v. The values m_v may differ across views. In many anchor-graph constructions, $\boldsymbol{w}_i^{(v)}$ is nonnegative and approximately simplex-valued, but we only require bounded vectors for the core derivations. The task is to infer a shared cluster label $z_i \in \{1, \ldots, K\}$ for every sample without aligning anchors and without requiring $m_1 = \cdots = m_V$.

We use the following notation convention throughout the paper and appendices: scalars are italic lowercase, vectors are bold lowercase (e.g., $\boldsymbol{\pi}$, $\boldsymbol{w}_i^{(v)}$, $\boldsymbol{c}_k^{(v)}$), matrices are bold uppercase (e.g., $\mathbf{W}^{(v)}$, $\mathbf{C}^{(v)}$, $\boldsymbol{\Phi}$), and distribution names are written in upright roman form (e.g., Dir, Cat, Gamma, and $\mathcal{N}$). In particular, the Dirichlet prior uses a vector concentration parameter $\boldsymbol{\alpha}_0 \in \mathbb{R}_{>0}^K$ unless a symmetric special case is stated explicitly.

3.2 Latent-label generation

We use a shared categorical latent variable to encode sample identity at the cluster level:

$$\boldsymbol{\pi} \sim \mathrm{Dir}(\boldsymbol{\alpha}_0), \tag{1}$$

$$z_i \mid \boldsymbol{\pi} \sim \mathrm{Cat}(\boldsymbol{\pi}), \qquad i = 1, \ldots, n. \tag{2}$$

Here $\boldsymbol{\pi} \in \Delta^{K-1}$ is the global mixture proportion and z_i is the cluster label of sample i. We assume $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0K})$ has strictly positive entries.

3.3 View-specific Gaussian observations

For each view v, we introduce a free deterministic view-specific prototype matrix

$$\mathbf{C}^{(v)} = [\boldsymbol{c}_1^{(v)}, \ldots, \boldsymbol{c}_K^{(v)}]^\top \in \mathbb{R}^{K \times m_v}, \tag{3}$$

where $\boldsymbol{c}_k^{(v)}$ is the prototype of cluster k in the anchor coordinate system of view v. The term free emphasizes two aspects of the construction. Across views, the prototype matrices are not tied by shared anchors, correspondence constraints, common bases, or cross-view regularizers. Within each view, the prototypes are treated as deterministic parameters optimized directly in the M-step, without imposing additional structural constraints. This choice yields a simple closed-form update, while the resulting prototypes still inherit basic geometric properties from the anchor vectors, as established in Section 4. The observation model is

$$\tau_v \sim \mathrm{Gamma}(a_0, b_0), \qquad v = 1, \ldots, V, \tag{4}$$

$$\boldsymbol{w}_i^{(v)} \mid z_i = k, \tau_v, \mathbf{C}^{(v)} \sim \mathcal{N}\Big(\boldsymbol{c}_k^{(v)},\, \tau_v^{-1}\mathbf{I}_{m_v}\Big). \tag{5}$$

We use the shape–rate parameterization $\mathrm{Gamma}(a, b)$. The scalar precision τ_v measures how tightly view v concentrates around its prototypes. Larger posterior τ_v means that view v is better explained by the current clustering and therefore exerts stronger influence on label posteriors. No anchor alignment appears anywhere in the model. The only shared latent quantity is the cluster label z_i.

3.4 Joint distribution

Let $\mathcal{W} = \{\boldsymbol{w}_i^{(v)} : i = 1, \ldots, n,\ v = 1, \ldots, V\}$, $\boldsymbol{z} = (z_1, \ldots, z_n)$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_V)$. The joint distribution is

$$p(\mathcal{W}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \{\mathbf{C}^{(v)}\}) = p(\boldsymbol{\pi}) \prod_{i=1}^{n} p(z_i \mid \boldsymbol{\pi}) \prod_{v=1}^{V} p(\tau_v) \prod_{v=1}^{V} \prod_{i=1}^{n} p\big(\boldsymbol{w}_i^{(v)} \mid z_i, \tau_v, \mathbf{C}^{(v)}\big). \tag{6}$$

Details of each distribution and the complete log-joint distribution are provided in Appendix A.1.

3.5 Variational Inference

3.5.1 Evidence lower bound

For fixed prototype parameters $\{\mathbf{C}^{(v)}\}_{v=1}^V$, the marginal likelihood, or evidence, is

$$p(\mathcal{W} \mid \{\mathbf{C}^{(v)}\}_{v=1}^V) = \sum_{\mathbf{z}} \int p(\mathcal{W}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \{\mathbf{C}^{(v)}\}_{v=1}^V)\, d\boldsymbol{\pi}\, d\boldsymbol{\tau}. \tag{7}$$

Directly maximizing this evidence is difficult because it requires summing over all latent label configurations and integrating over $\boldsymbol{\pi}$ and $\boldsymbol{\tau}$. We therefore introduce a variational distribution

$q(\boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\tau})$ and have

$$\log p(\mathcal{W} \mid \{\mathbf{C}^{(v)}\}) = \log \sum_{\mathbf{z}} \int q(\boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\tau}) \frac{p(\mathcal{W}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \{\mathbf{C}^{(v)}\})}{q(\boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\tau})} \, d\boldsymbol{\pi} \, d\boldsymbol{\tau}$$
$$\geq \sum_{\mathbf{z}} \int q(\boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\tau}) \log \frac{p(\mathcal{W}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \{\mathbf{C}^{(v)}\})}{q(\boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\tau})} \, d\boldsymbol{\pi} \, d\boldsymbol{\tau}. \tag{8}$$

Hence, the log evidence lower bound is

$$\mathcal{L}\big(q, \{\mathbf{C}^{(v)}\}\big) = \mathbb{E}_q\Big[\log p(\mathcal{W}, \boldsymbol{z}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \{\mathbf{C}^{(v)}\})\Big] - \mathbb{E}_q[\log q(\boldsymbol{\pi}, \boldsymbol{z}, \boldsymbol{\tau})], \tag{9}$$

whose explicit expression is provided in Appendix A.2.

3.5.2 Design of posterior distribution

For a product family $q(\theta) = \prod_j q_j(\theta_j)$, Coordinate-Ascent Variational Inference (CAVI) gives the optimal factor update

$$\log q_j^\star(\theta_j) = \mathbb{E}_{q_{-j}}[\log p(\mathcal{W}, \theta \mid \{\mathbf{C}^{(v)}\})] + \text{const}, \tag{10}$$

where q_{-j} denotes all factors except the jth. This identity motivates choosing blocks that preserve conjugacy and exploit conditional independence across samples and views.

We therefore use the mean-field family

$$q(\boldsymbol{\pi}, \boldsymbol{z}, \boldsymbol{\tau}) = q(\boldsymbol{\pi}) \prod_{i=1}^{n} q(z_i) \prod_{v=1}^{V} q(\tau_v), \tag{11}$$

The factorization in (11) is the mean-field approximation: it assumes posterior independence among $\boldsymbol{\pi}$, the sample labels $\{z_i\}_{i=1}^n$, and the view precisions $\{\tau_v\}_{v=1}^V$. However, the parametric forms of the coordinate-optimal factors are not chosen arbitrarily. They follow from the generic CAVI identity. Because the Dirichlet prior is conjugate to the categorical latent labels, the coordinate-optimal factor for $\boldsymbol{\pi}$ remains Dirichlet. Similarly, because the Gamma prior is conjugate to the Gaussian precision, the coordinate-optimal factor for each τ_v remains Gamma. The prototype matrices $\{\mathbf{C}^{(v)}\}$ are treated as deterministic parameters updated in an M-step rather than random variables with their own variational factors. This choice keeps the algorithm in the coordinate-ascent variational EM regime and preserves simple weighted-average prototype updates.

Coordinate-optimal form for $q(z_i)$. Since $z_i \in \{1, \ldots, K\}$ is a K-state discrete latent variable, an arbitrary variational factor over z_i can be parameterized by a probability vector $\boldsymbol{\phi}_i \in \Delta^{K-1}$. The quantities ϕ_{ik} are the variational responsibilities, representing the posterior probability under q_i that sample i belongs to cluster k, i.e., $q_i(z_i = k) = \phi_{ik}$. Therefore

$$q_i(z_i) = \text{Cat}(\boldsymbol{\phi}_i), \qquad \sum_{k=1}^{K} \phi_{ik} = 1, \qquad \phi_{ik} \geq 0. \tag{12}$$

Coordinate-optimal form of $q(\boldsymbol{\pi})$. To derive $q^\star(\boldsymbol{\pi})$, we retain only the $\boldsymbol{\pi}$-dependent terms in the complete-data log joint and apply the CAVI identity, giving

$$\log q^\star(\boldsymbol{\pi}) = \mathbb{E}_{-\boldsymbol{\pi}}\left[\log p(\boldsymbol{\pi}) + \sum_{i=1}^{n} \log p(z_i \mid \boldsymbol{\pi})\right] + \text{const}$$
$$= \mathbb{E}_{-\boldsymbol{\pi}}\left[\sum_{k=1}^{K} (\alpha_{0k} - 1) \log \pi_k + \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log \pi_k\right] + \text{const}$$
$$= \sum_{k=1}^{K} \left(\alpha_{0k} - 1 + \sum_{i=1}^{n} \phi_{ik}\right) \log \pi_k + \text{const}. \tag{13}$$

where $\mathbb{E}_{q(z_i)}[z_{ik}] = q(z_i = k) = \phi_{ik}$. This is the log density of a Dirichlet distribution. Therefore

$$q^\star(\boldsymbol{\pi}) = \text{Dir}(\tilde{\boldsymbol{\alpha}}), \qquad \tilde{\alpha}_k = \alpha_{0k} + \sum_{i=1}^{n} \phi_{ik}. \tag{14}$$

Coordinate-optimal form for $q_v(\tau_v)$. To derive $q_v^\star(\tau_v)$, we retain only the τ_v-dependent terms in the complete-data log joint and apply the CAVI identity, giving

$$\begin{aligned}\log q_v^\star(\tau_v) &= \mathbb{E}_{-\tau_v}\left[\log p(\tau_v) + \sum_{i=1}^{n} \log p(\boldsymbol{w}_i^{(v)} \mid z_i, \tau_v, \mathbf{C}^{(v)})\right] + \text{const} \\ &= (a_0 - 1)\log\tau_v - b_0\tau_v \\ &\quad + \sum_{i=1}^{n}\sum_{k=1}^{K} \mathbb{E}_{q_i}[z_{ik}]\left[\frac{m_v}{2}\log\tau_v - \frac{\tau_v}{2}\left\|\boldsymbol{w}_i^{(v)} - \mathbf{c}_k^{(v)}\right\|_2^2\right] + \text{const} \\ &= \left(a_0 + \frac{nm_v}{2} - 1\right)\log\tau_v - \left(b_0 + \frac{1}{2}\text{SSE}_v\right)\tau_v + \text{const}. \end{aligned} \tag{15}$$

where $\text{SSE}_v = \sum_{i=1}^{n}\sum_{k=1}^{K}\phi_{ik}\left\|\boldsymbol{w}_i^{(v)} - \mathbf{c}_k^{(v)}\right\|_2^2$ is the view-specific expected residual sum of squares and $\frac{m_v}{2}\sum_{i=1}^{n}\sum_{k=1}^{K}\phi_{ik} = \frac{m_v}{2}\sum_{i=1}^{n}1 = \frac{nm_v}{2}$. This is the log density of a Gamma distribution under the shape–rate parameterization. Therefore,

$$q_v^\star(\tau_v) = \text{Gamma}(\tilde{a}_v, \tilde{b}_v), \qquad \tilde{a}_v = a_0 + \frac{nm_v}{2}, \qquad \tilde{b}_v = b_0 + \frac{1}{2}\text{SSE}_v. \tag{16}$$

3.5.3 Closed-form E-step

Update of $q(\boldsymbol{\pi})$. By the standard log-moment identity of the Dirichlet distribution, we have

$$\mathbb{E}_q[\log\pi_k] = \psi(\tilde{\alpha}_k) - \psi\left(\sum_{h=1}^{K}\tilde{\alpha}_h\right), \tag{17}$$

where $\psi(\cdot)$ denotes the digamma function. A derivation is provided in Appendix A.3.

Update of $q(\tau_v)$. For each view precision, the required posterior moments are

$$\bar{\tau}_v := \mathbb{E}_q[\tau_v] = \frac{\tilde{a}_v}{\tilde{b}_v}, \qquad \mathbb{E}_q[\log\tau_v] = \psi(\tilde{a}_v) - \log\tilde{b}_v. \tag{18}$$

These are standard Gamma moment identities; a derivation is provided in Appendix A.4. The first identity follows directly from the mean of a Gamma distribution under the shape–rate parameterization. This makes the view-weighting effect transparent: views with smaller residual dispersion automatically obtain larger posterior precision.

Update of $q_i(z_i)$. We define the reduced log-score and distance

$$\xi_{ik} = \mathbb{E}_q[\log\pi_k] - \frac{1}{2}\sum_{v=1}^{V}\bar{\tau}_v D_{ik}^{(v)}, \qquad D_{ik}^{(v)} = \left\|\mathbf{w}_i^{(v)} - \mathbf{c}_k^{(v)}\right\|_2^2, \tag{19}$$

resulting in the following $\boldsymbol{\phi}_i$-dependent ELBO terms

$$\mathcal{L}_i(\boldsymbol{\phi}_i) = \sum_{k=1}^{K}\phi_{ik}\xi_{ik} - \sum_{k=1}^{K}\phi_{ik}\log\phi_{ik} + \text{const}. \tag{20}$$

Therefore the solution of $\boldsymbol{\phi}_i$-subproblem obeys Proposition 3.1.

Proposition 3.1 (Responsibility block as entropy-regularized assignment; proof in App. A.5)**.** *For fixed $q(\boldsymbol{\pi})$, $q(\boldsymbol{\tau})$, and $\{\mathbf{C}^{(v)}\}$, the optimal responsibility vector for sample i is the unique solution of*

$$\max_{\boldsymbol{\phi}_i \in \Delta^{K-1}}\left\{\sum_{k=1}^{K}\phi_{ik}\xi_{ik} - \sum_{k=1}^{K}\phi_{ik}\log\phi_{ik}\right\}. \tag{21}$$

Its unique maximizer is exactly the softmax of $\boldsymbol{\xi}_i$, i.e.,

$$\phi_{ik} = \frac{\exp(\xi_{ik})}{\sum_{\ell=1}^{K}\exp(\xi_{i\ell})}. \tag{22}$$

In particular, $\phi_{ik} > 0$ for every i and k in exact arithmetic, so the prototype update never encounters a truly empty cluster.

3.6 M-step: view-specific prototype updates

Fixing q, the ELBO terms involving $\{\mathbf{C}^{(v)}\}$ reduce to the weighted least-squares objective

$$\mathcal{L}\big(\{\mathbf{C}^{(v)}\}\big) = -\frac{1}{2}\sum_{v=1}^{V}\bar{\tau}_v \sum_{i=1}^{n}\sum_{k=1}^{K}\phi_{ik}\|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}\|_2^2. \tag{23}$$

This decouples completely across views and clusters. For each (v, k) we solve

$$\min_{\boldsymbol{c}_k^{(v)}\in\mathbb{R}^{m_v}} \sum_{i=1}^{n}\phi_{ik}\|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}\|_2^2, \tag{24}$$

whose solution is the weighted average, recorded by Proposition 3.2.

Proposition 3.2 (Closed-form view-specific prototype update; proof in App. A.6)**.** *For fixed q, the optimal prototype of cluster k in view v is*

$$\boldsymbol{c}_k^{(v)} = \frac{\sum_{i=1}^{n}\phi_{ik}\boldsymbol{w}_i^{(v)}}{\sum_{i=1}^{n}\phi_{ik}}, \tag{25}$$

where the denominator is strictly positive in exact arithmetic by Proposition 3.1.

Putting these closed-form updates together, we term the resulting method Probabilistic Multi-View Anchor Graph Clustering (PMVAGC), with its algorithmic procedure summarized in Appendix C.

4 Theoretical Analysis

Prototype geometry. Although the prototypes are free from cross-view correspondence constraints and are not explicitly constrained to be probability vectors, their M-step updates still inherit the within-view geometry of the observed anchor vectors.

Proposition 4.1 (Convex hull and simplex closure; proof in App. A.7)**.** *For any fixed view v and cluster k, the update in Eq.* (25) *lies in the convex hull of $\{\boldsymbol{w}_i^{(v)}\}_{i=1}^n$. Consequently, if every anchor vector in view v satisfies $\boldsymbol{w}_i^{(v)} \in \Delta^{m_v-1}$, then every prototype updated by Eq.* (25) *also satisfies $\boldsymbol{c}_k^{(v)} \in \Delta^{m_v-1}$.*

Permutation invariance. A direct mathematical expression of the alignment-free viewpoint is invariance to within-view anchor reordering.

Proposition 4.2 (View-wise anchor permutation invariance; proof in App. A.8)**.** *Fix a view v and any permutation matrix $\mathbf{P} \in \mathbb{R}^{m_v\times m_v}$. Replace every $\boldsymbol{w}_i^{(v)}$ by $\mathbf{P}\boldsymbol{w}_i^{(v)}$ and every prototype $\boldsymbol{c}_k^{(v)}$ by $\mathbf{P}\boldsymbol{c}_k^{(v)}$. Then all distances $D_{ik}^{(v)}$ are unchanged, and therefore the ELBO, every variational update, and the final labels are unchanged. Equivalently, the algorithm is equivariant to arbitrary reorderings of anchor coordinates within each view.*

Adaptive view weighting through posterior precisions. The model compares views through a *regularized inverse residual variance per anchor dimension*. When a_0 and b_0 are weak relative to the data term, $\bar{\tau}_v$ is approximately the inverse average squared residual per anchor coordinate.

Proposition 4.3 (Residual-based adaptive weighting; proof in App. A.9)**.** *For fixed n, m_v, a_0, b_0, the posterior precision $\bar{\tau}_v$ is strictly decreasing in SE_v. Consequently, among views with comparable dimensionality, lower reconstruction error yields larger influence in the responsibility update.*

Convergence and stationarity. For compactness, we write $\mathbf{C} = [\mathbf{C}^{(1)}, \ldots, \mathbf{C}^{(V)}]$ for the collection of all view-specific prototype matrices and state the convergence analysis using the reduced state $(\boldsymbol{\Phi}, \bar{\boldsymbol{\tau}}, \mathbf{C})$. The variables $\boldsymbol{\pi}$ and $\boldsymbol{\tau}$ are still represented by their variational factors; $\bar{\tau}_v$ denotes only the posterior mean of the Gamma factor for τ_v, and the Dirichlet factor for $\boldsymbol{\pi}$ is determined by $\boldsymbol{\Phi}$. Let the algorithmic state at iteration t be $(\boldsymbol{\Phi}^{(t)}, \bar{\boldsymbol{\tau}}^{(t)}, \mathbf{C}^{(t)})$.

Theorem 4.4 (Monotone convergence, uniform interior responsibilities, vanishing updates, and limit-point stationarity)**.** *Assume that for every view v, the data set $\mathcal{W}_v = \{\boldsymbol{w}_i^{(v)}\}_{i=1}^n$ is finite, that $\alpha_{0k} > 0$ for all k, and that $a_0 > 0$ and $b_0 > 0$. Then the iterates produced by Algorithm 1 satisfy the following properties.*

(i) The ELBO sequence $\{\mathcal{L}^{(t)}\}$ is nondecreasing and convergent.

(ii) There exists $\varepsilon > 0$ such that $\phi_{ik}^{(t)} \geq \varepsilon$ for all i, k, t. Hence every prototype update in Eq. (25) is well defined without any exact empty cluster.

(iii) The successive differences vanish:

$$\|\mathbf{\Phi}^{(t+1)} - \mathbf{\Phi}^{(t)}\|_F \to 0, \qquad \|\mathbf{C}^{(t+1)} - \mathbf{C}^{(t)}\|_F \to 0, \qquad \|\bar{\boldsymbol{\tau}}^{(t+1)} - \bar{\boldsymbol{\tau}}^{(t)}\|_2 \to 0.$$

(iv) Every accumulation point of $\{(\mathbf{\Phi}^{(t)}, \bar{\boldsymbol{\tau}}^{(t)}, \mathbf{C}^{(t)})\}$ is a fixed point of the full cyclic update map and therefore a coordinatewise stationary point of the variational EM objective.

The proof is fully constructive in Appendix B. The key ingredients are: (a) prototypes remain in compact convex hulls by Proposition 4.1; (b) the responsibility block is strictly interior because of the entropy term; (c) the responsibility-update gain is an exact sum of KL divergences; and (d) the prototype-update gain is a strictly positive quadratic form controlling the step norm. This yields not just monotonicity but vanishing increments and a fixed-point characterization of all limit points.

Time complexity. In $q_\pi(\boldsymbol{\pi})$-update step, the Dirichlet parameter update and subsequent expectation calculation require $O(nK)$ and $O(K)$, respectively. The distance computation between all samples and prototypes from all views takes $O(nK\sum_{v=1}^{V} m_v)$. Once the distances are already available, the $q_v(\tau_v)$-update step and $q_i(z_i)$-update step require $O(nKV)$ and $O(nKV + nK)$, respectively. In prototype-update step, computing K prototypes over all views costs $O(nK\sum_{v=1}^{V} m_v)$. If the ELBO is evaluated at every iteration, $O(nKV + nK + K + V)$ is required. Thus one outer iteration has time complexity $O(nK\sum_{v=1}^{V} m_v + nKV)$, commonly reported as $O(nK\sum_{v=1}^{V} m_v)$.

Memory complexity. The input anchor graphs, prototype matrices and responsibility matrix Φ require $O(n\sum_{v=1}^{V} m_v)$, $O(K\sum_{v=1}^{V} m_v)$ and $O(nK)$. The variational parameters $\tilde{\boldsymbol{\alpha}}$, $\{(\tilde{a}_v, \tilde{b}_v)\}_{v=1}^{V}$, and $\bar{\boldsymbol{\tau}}$ require $O(K + V)$. If all view-specific distances $D_{ik}^{(v)}$ are stored, $O(nKV)$ is required. Therefore, the memory complexity is $O(n\sum_{v=1}^{V} m_v + K\sum_{v=1}^{V} m_v + nKV)$.

5 Experiments

5.1 Experimental setup

In our experiments, 10 publicly available datasets (MSRC, ORL, 100Leaves, NGs, MNIST, YTF10, YTF20, YTF50, CIFAR10, CIFAR100; details in Appendix D.1), 13 representative algorithms (LMVSC [Kang et al., 2020], FPMVS [Wang et al., 2022b], EMVGC [Wen et al., 2023], SFMC [Li et al., 2022], FDAGF [Zhang et al., 2023], FMVACC [Wang et al., 2022a], OMSC [Chen et al., 2022], AEVC [Liu et al., 2024], ESTMC [Ji and Feng, 2025], MVS4C [Lu et al., 2025a], RTGD [Zou et al., 2025], TMC [Wang et al., 2026], SLWSE [Tang et al., 2026]; details in Appendix D.2) and 3 well-known metrics (Accuracy (ACC), Normalized Mutual Information (NMI), Adjusted Rand Index (ARI)) are used for empirical evaluation. The input anchor graphs of our model are generated by the basic anchor subspace model $\min_{\mathbf{A},\mathbf{W}} \|\mathbf{X} - \mathbf{W}\mathbf{A}\|_F^2$, $\mathbf{W}\mathbf{1} = \mathbf{1}, \mathbf{W} \geq 0$, which is consistent with our most relevant algorithms, LMVSC, FMVACC, and AEVC. In particular, we denote the variant of PMVAGC with heterogeneous anchor cardinalities by PMVAGC-H. For all compared methods, hyperparameters other than the anchor number are searched according to the original papers. For the anchor number, we use a unified search range $m \in \{c, 2c, 3c\}$, where c is the number of clusters. We report the best mean performance and standard deviation over five runs. All experiments are conducted on a machine with an Intel i7-14700 CPU and 64GB RAM.

5.2 Experimental results

Clustering performance. As shown in Table 1, PMVAGC achieves the best results on most datasets and metrics, and ranks second on the remaining cases. This demonstrates that our algorithm not only liberates anchors from cross-view correspondence but also delivers competitive performance. Further experiments, recorded in Table 2, were conducted to explore the algorithm's inherent ability to allow heterogeneous anchor cardinalities. PMVAGC-H achieves gains on datasets other than CIFAR100, with particularly pronounced improvements on some datasets and metrics. This indicates

that different combinations of anchor numbers across views help to further unlock the model's potential. In summary, our model is well-suited for the aggregation of single-view models, while also highlighting the importance of avoiding alignment and allowing unequal anchor numbers.

View weights and controlled diagnostics. Figure 1 visualizes the normalized view weights derived from posterior precisions, while Table 3 compares PMVAGC (Adapt.) with two controls. The Equal variant uses uniform view contributions in the responsibility update with all other settings unchanged; the Single-View Oracle (SVO) selects the best single view for each dataset and metric retrospectively using ground-truth labels. Adapt. improves 22 of 30 dataset–metric cases over Equal and matches or improves 27, with mean ACC, NMI, and ARI gains of 1.38, 1.89, and 2.12 percentage points, respectively. Against SVO, it improves 28 of 30 cases and matches or improves 29, with corresponding gains of 11.03, 13.62, and 18.33 percentage points. These results support both adaptive posterior precision weighting and complementary multi-view integration.

Runtime. As shown in Figure 2, our model's runtime is generally below the median, and often second only to MVS4C and within the same order of magnitude, while achieving substantially better clustering performance than the faster alternatives. Furthermore, for datasets like MNIST with many samples and few classes, our runtime is still quite advantageous. This setting is common in many practical clustering scenarios. In conclusion, our current runtime can be regarded as practical.

Convergence. Figure 3 shows the trend of PMVAGC's ELBO, which first increases and then tends to level off as it iterates, empirically validating the convergence behavior of our model.

Table 1: Clustering performance comparison on all datasets.

ACC ↑

Method	MSRC	ORL	100Leaves	NGs	MNIST	YTF10	YTF20	YTF50	CIFAR10	CIFAR100
LMVSC	63.33±0.00	74.75±0.00	55.88±0.00	32.60±0.00	98.88±0.00	71.27±0.00	65.10±0.00	66.89±0.00	88.98±0.00	76.81±0.00
FPMVS	86.19±0.00	56.75±0.00	37.56±0.00	73.80±0.00	98.84±0.00	74.96±0.00	68.66±0.00	OOM	98.99±0.00	OOM
EMVGC	39.05±0.00	74.25±0.00	72.19±0.00	27.80±0.00	98.78±0.00	69.35±0.00	66.67±0.00	69.09±0.00	99.01±0.00	91.61±0.00
SFMC	79.05±3.97	74.25±2.32	69.63±0.97	79.60±7.67	98.58±0.15	65.19±2.72	57.99±2.19	OOM	95.72±5.42	**99.97±0.00**
FDAGF	87.62±0.00	72.75±0.00	63.44±0.00	70.40±0.00	98.87±0.00	73.04±0.00	65.62±0.00	67.87±0.00	98.72±0.00	85.42±0.00
FMVACC	82.86±0.00	79.25±0.00	41.06±0.00	72.40±0.00	98.89±0.00	76.56±0.00	71.36±0.00	71.08±0.00	99.03±0.00	91.70±0.00
OMSC	83.81±0.00	59.75±0.00	38.00±0.00	77.80±0.00	98.91±0.00	79.22±0.00	69.86±0.00	OOM	OOM	OOM
AEVC	85.24±0.00	77.00±0.00	71.31±0.00	93.80±0.00	98.75±0.00	75.53±0.00	70.34±0.00	71.91±0.00	96.27±0.00	92.18±0.00
ESTMC	87.62±0.00	78.50±0.00	66.81±0.00	84.20±0.00	98.44±0.00	77.48±0.00	**75.30±0.00**	73.62±0.00	**99.34±0.00**	93.50±0.00
MVS4C	75.40±1.70	62.83±0.76	62.13±1.86	77.73±1.44	98.29±0.00	73.57±4.44	65.33±4.30	68.89±0.73	97.53±2.96	96.43±0.70
RTGD	57.14±0.00	77.75±0.00	48.19±0.00	95.00±0.00	96.59±0.00	77.92±0.00	66.68±0.00	64.25±0.00	85.79±0.00	76.67±0.00
TMC	83.52±2.90	69.95±3.22	67.48±2.02	88.76±0.55	98.75±0.00	78.77±1.40	71.38±3.42	71.50±1.64	98.97±0.00	91.61±1.67
SLWSE	77.14±0.00	66.25±0.00	32.25±0.00	68.40±0.00	98.76±0.00	81.32±0.00	67.98±0.00	67.60±0.00	98.98±0.00	97.50±0.00
PMVAGC	**91.75±0.95**	**80.42±2.02**	**72.21±2.39**	**95.80±0.00**	**98.92±0.00**	**83.05±0.95**	73.68±3.14	**74.98±1.15**	99.17±0.00	**99.97±1.51**

NMI ↑

Method	MSRC	ORL	100Leaves	NGs	MNIST	YTF10	YTF20	YTF50	CIFAR10	CIFAR100
LMVSC	54.53±0.00	88.33±0.00	77.59±0.00	10.11±0.00	96.63±0.00	76.89±0.00	74.76±0.00	80.59±0.00	78.91±0.00	93.67±0.00
FPMVS	78.09±0.00	77.31±0.00	64.65±0.00	58.07±0.00	96.49±0.00	78.69±0.00	75.15±0.00	OOM	97.30±0.00	OOM
EMVGC	31.29±0.00	87.47±0.00	86.37±0.00	08.15±0.00	96.32±0.00	69.76±0.00	76.95±0.00	83.22±0.00	97.40±0.00	98.15±0.00
SFMC	76.50±2.74	85.17±1.14	82.58±0.60	69.87±6.11	96.15±0.26	67.68±4.49	62.52±1.99	OOM	95.23±2.89	**99.96±0.00**
FDAGF	79.02±0.00	87.97±0.00	83.38±0.00	45.60±0.00	96.57±0.00	77.84±0.00	77.31±0.00	80.56±0.00	96.58±0.00	96.72±0.00
FMVACC	72.00±0.00	89.58±0.00	67.48±0.00	51.19±0.00	96.56±0.00	80.19±0.00	78.36±0.00	82.22±0.00	97.37±0.00	98.03±0.00
OMSC	72.95±0.00	78.74±0.00	65.17±0.00	55.38±0.00	96.67±0.00	81.07±0.00	76.18±0.00	OOM	OOM	OOM
AEVC	75.83±0.00	87.31±0.00	86.35±0.00	84.30±0.00	96.22±0.00	78.17±0.00	76.24±0.00	82.82±0.00	92.17±0.00	98.14±0.00
ESTMC	77.80±0.00	89.45±0.00	83.84±0.00	66.18±0.00	95.73±0.00	76.87±0.00	78.47±0.00	83.99±0.00	**98.19±0.00**	98.46±0.00
MVS4C	62.34±2.06	78.05±0.65	78.95±0.51	56.01±2.56	95.36±0.00	75.47±2.23	71.66±2.59	77.10±0.11	96.20±3.09	98.39±0.06
RTGD	50.26±0.00	87.78±0.00	70.23±0.00	86.97±0.00	92.63±0.00	82.18±0.00	77.37±0.00	82.27±0.00	93.81±0.00	90.94±0.00
TMC	73.10±2.52	84.75±1.32	84.44±0.86	74.83±0.85	96.38±0.00	81.39±1.16	78.74±1.11	82.89±0.59	97.13±0.00	98.63±0.26
SLWSE	74.86±0.00	83.85±0.00	63.29±0.00	54.89±0.00	96.31±0.00	81.91±0.00	77.23±0.00	83.26±0.00	97.42±0.00	99.56±0.00
PMVAGC	**85.39±1.03**	**90.25±0.41**	**88.58±0.92**	**88.30±0.00**	**96.74±0.00**	**82.51±0.10**	**80.25±0.73**	**84.00±0.27**	97.70±0.00	**99.96±0.35**

ARI ↑

Method	MSRC	ORL	100Leaves	NGs	MNIST	YTF10	YTF20	YTF50	CIFAR10	CIFAR100
LMVSC	41.36±0.00	66.87±0.00	44.72±0.00	04.23±0.00	97.50±0.00	64.06±0.00	52.78±0.00	55.20±0.00	76.93±0.00	73.38±0.00
FPMVS	72.32±0.00	46.07±0.00	21.39±0.00	56.61±0.00	97.42±0.00	67.06±0.00	61.16±0.00	OOM	97.78±0.00	OOM
EMVGC	15.15±0.00	66.84±0.00	63.02±0.00	01.36±0.00	97.31±0.00	50.55±0.00	55.13±0.00	61.73±0.00	97.83±0.00	92.77±0.00
SFMC	64.01±3.65	45.63±3.34	41.12±3.58	66.75±7.92	96.90±0.32	42.88±7.08	23.37±2.62	OOM	94.49±5.20	**99.94±0.01**
FDAGF	74.44±0.00	67.16±0.00	53.15±0.00	37.09±0.00	97.49±0.00	65.42±0.00	53.78±0.00	56.57±0.00	97.20±0.00	86.56±0.00
FMVACC	63.46±0.00	72.74±0.00	29.43±0.00	41.28±0.00	97.53±0.00	71.25±0.00	59.40±0.00	61.95±0.00	97.87±0.00	92.05±0.00
OMSC	68.51±0.00	48.15±0.00	23.32±0.00	53.70±0.00	97.56±0.00	72.11±0.00	61.22±0.00	OOM	OOM	OOM
AEVC	70.45±0.00	67.37±0.00	61.41±0.00	85.23±0.00	97.25±0.00	68.63±0.00	60.53±0.00	61.69±0.00	92.18±0.00	93.20±0.00
ESTMC	75.21±0.00	70.92±0.00	56.78±0.00	64.38±0.00	96.59±0.00	66.05±0.00	64.59±0.00	62.35±0.00	**98.55±0.00**	94.20±0.00
MVS4C	55.16±2.73	49.68±1.30	47.11±1.63	53.63±3.03	96.28±0.00	64.98±5.55	55.99±4.38	60.80±0.50	95.51±4.91	95.65±0.18
RTGD	38.05±0.00	69.87±0.00	31.19±0.00	87.88±0.00	92.79±0.00	69.77±0.00	60.99±0.00	56.53±0.00	86.97±0.00	55.55±0.00
TMC	66.79±3.66	59.63±3.66	57.38±2.00	74.84±1.08	97.26±0.00	71.62±2.27	59.33±5.58	60.08±2.51	97.74±0.00	92.81±1.37
SLWSE	65.80±0.00	56.33±0.00	20.47±0.00	48.49±0.00	97.26±0.00	72.15±0.00	60.92±0.00	59.75±0.00	97.78±0.00	97.69±0.00
PMVAGC	**81.08±1.47**	**73.47±1.00**	**64.81±2.98**	**89.63±0.00**	**97.61±0.00**	**74.93±0.38**	**65.63±0.45**	**62.66±0.74**	98.18±0.00	**99.94±1.36**

Note: Values are percentages. OOM means out of memory. PMVAGC uses the same number of anchors for all views.

Table 2: Effect of heterogeneous anchor cardinalities across views.

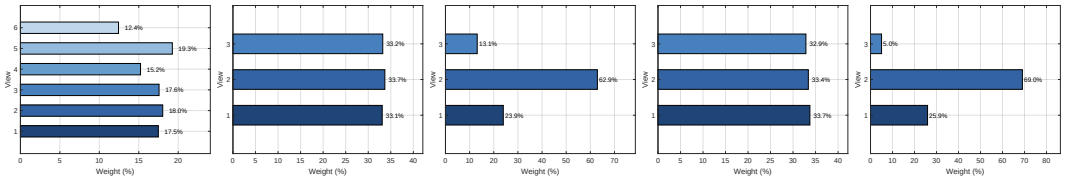
Dataset	ACC ↑			NMI ↑			ARI ↑		
	PMVAGC	PMVAGC-H	Gain	PMVAGC	PMVAGC-H	Gain	PMVAGC	PMVAGC-H	Gain
MSRC	91.75	94.29	+2.54	85.39	89.25	+3.86	81.08	86.59	+5.51
ORL	80.42	86.50	+6.08	90.25	92.37	+2.12	73.47	80.27	+6.80
100Leaves	72.21	74.63	+2.42	88.58	89.89	+1.31	64.81	67.48	+2.67
NGs	95.80	96.40	+0.60	88.30	89.63	+1.33	89.63	91.09	+1.46
MNIST	98.92	99.09	+0.17	96.74	97.19	+0.45	97.61	97.98	+0.37
YTF10	83.05	88.92	+5.87	82.51	86.58	+4.07	74.93	79.44	+4.51
YTF20	73.68	80.61	+6.93	80.25	81.99	+1.74	65.63	66.97	+1.34
YTF50	74.98	79.99	+5.01	84.00	85.44	+1.44	62.66	63.28	+0.62
CIFAR10	99.17	99.33	+0.16	97.70	98.15	+0.45	98.18	98.53	+0.35
CIFAR100	99.97	99.97	+0.00	99.96	99.96	+0.00	99.94	99.94	+0.00

Note: Values are percentages. PMVAGC uses the same number of anchors for all views, whereas PMVAGC-H allows view-specific anchor numbers. Gain denotes PMVAGC-H minus PMVAGC in percentage points.

Table 3: Controlled diagnostics of adaptive view weighting and multi-view integration.

Dataset	ACC ↑					NMI ↑					ARI ↑				
	SVO	Equal	Adapt.	Δ_{SVO}	Δ_{Eq}	SVO	Equal	Adapt.	Δ_{SVO}	Δ_{Eq}	SVO	Equal	Adapt.	Δ_{SVO}	Δ_{Eq}
MSRC	67.14	89.52	**91.75**	+24.61	+02.23	55.96	82.50	**85.39**	+29.43	+02.89	46.12	76.92	**81.08**	+34.96	+04.16
ORL	76.25	78.50	**80.42**	+04.17	+01.92	88.46	88.88	**90.25**	+01.79	+01.37	66.72	70.53	**73.47**	+06.75	+02.94
100Leaves	51.06	68.63	**72.21**	+21.15	+03.58	74.94	88.00	**88.58**	+13.64	+00.58	33.81	64.78	**64.81**	+31.00	+00.03
NGs	51.00	**95.80**	**95.80**	+44.80	+00.00	29.10	**88.30**	**88.30**	+59.20	+00.00	14.85	**89.63**	**89.63**	+74.78	+00.00
MNIST	98.91	95.18	**98.92**	+00.01	+03.74	96.65	89.73	**96.74**	+00.09	+07.01	**97.61**	89.99	**97.61**	+00.00	+07.62
YTF10	80.19	83.01	**83.05**	+02.86	+00.04	76.49	80.84	**82.51**	+06.02	+01.67	66.60	73.14	**74.93**	+08.33	+01.79
YTF20	**75.47**	74.02	73.68	–01.79	–00.34	77.14	79.22	**80.25**	+03.11	+01.03	60.90	63.74	**65.63**	+04.73	+01.89
YTF50	72.62	73.91	**74.98**	+02.36	+01.07	80.57	83.09	**84.00**	+03.43	+00.91	61.59	**63.21**	62.66	+01.07	–00.55
CIFAR10	89.82	97.59	**99.17**	+09.35	+01.58	78.87	94.29	**97.70**	+18.83	+03.41	79.04	94.81	**98.18**	+19.14	+03.37
CIFAR100	97.19	**99.97**	**99.97**	+02.78	+00.00	99.33	**99.96**	**99.96**	+00.63	+00.00	97.45	**99.95**	99.94	+02.49	–00.01

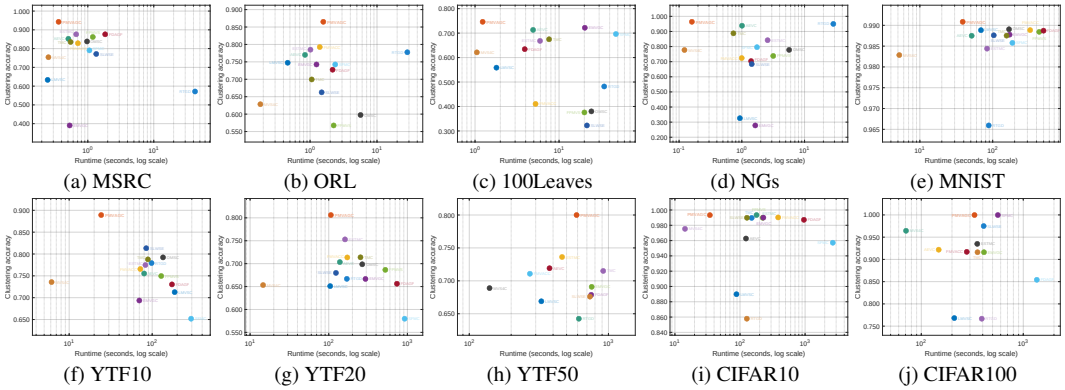
Note: Scores are percentages; bold marks the best score (ties included). Adapt. is full PMVAGC with adaptive view weights; Equal uses uniform view weights in responsibility updates. SVO (Single-View Oracle) retrospectively selects the best view per dataset and metric using ground-truth labels. $\Delta_{\mathrm{SVO}}/\Delta_{\mathrm{Eq}}$ are Adapt. minus SVO/Equal, respectively (percentage points).



(a) MSRC (b) ORL (c) 100Leaves (d) NGs (e) MNIST

(f) YTF10 (g) YTF20 (h) YTF50 (i) CIFAR10 (j) CIFAR100

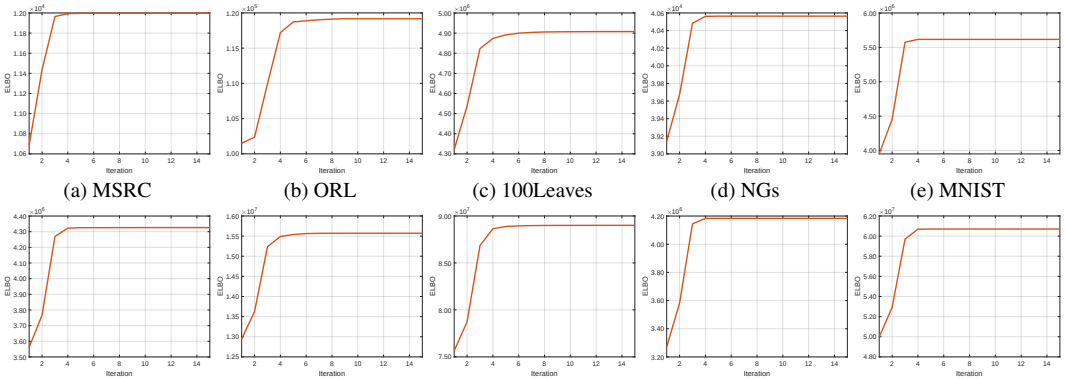
Figure 1: View-specific weight distributions of our algorithm on benchmark datasets.



(a) MSRC (b) ORL (c) 100Leaves (d) NGs (e) MNIST

(f) YTF10 (g) YTF20 (h) YTF50 (i) CIFAR10 (j) CIFAR100

Figure 2: Accuracy–runtime trade-offs of different methods on benchmark datasets.



(a) MSRC (b) ORL (c) 100Leaves (d) NGs (e) MNIST

(f) YTF10 (g) YTF20 (h) YTF50 (i) CIFAR10 (j) CIFAR100

Figure 3: ELBO curves of our algorithm on benchmark datasets.

6 Conclusion

We presented PMVAGC, an alignment-free probabilistic framework for multi-view anchor graph clustering. By coupling views only through shared latent cluster labels, PMVAGC avoids explicit anchor correspondence while allowing view-specific anchor spaces and anchor cardinalities. The resulting variational EM algorithm admits closed-form updates, adaptive view weighting, and convergence guarantees. Experiments demonstrate competitive clustering performance, practical runtime, and the benefit of heterogeneous anchor numbers across views.

Acknowledgments and Disclosure of Funding

We thank the anonymous reviewers for their valuable comments and suggestions. This work was supported by the University of Macau and University of Macau Development Foundation under Grant MYRG-GRG2024-00181-FST-UMDF. The authors declare no competing interests.

References

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Man-Sheng Chen, Chang-Dong Wang, Dong Huang, Jian-Huang Lai, and Philip S Yu. Efficient orthogonal multi-view subspace clustering. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 127–135, 2022.

Yawei Chen, Huibing Wang, Jinjia Peng, and Yang Wang. Anchor learning with potential cluster constraints for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39(15), pages 15939–15947, 2025.

Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38, 1977.

Jintian Ji and Songhe Feng. Anchors crash tensor: Efficient and scalable tensorial multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4): 2660–2675, 2025.

Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34(04), pages 4412–4419, 2020.

Zisen Kong, Dongxia Chang, Yiming Wang, Pengyuan Li, and Yao Zhao. Tensorial multi-view clustering via alternative rank minimization and inter-view alignment. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

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.

Suyuan Liu, Ke Liang, Zhibin Dong, Siwei Wang, Xihong Yang, Sihang Zhou, En Zhu, and Xinwang Liu. Learn from view correlation: An anchor enhancement strategy for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26151–26161, 2024.

Zhoumin Lu, Feiping Nie, Linru Ma, Rong Wang, and Xuelong Li. Simplifying scalable subspace clustering and its multi-view extension by anchor-to-sample kernel. *IEEE Transactions on Image Processing*, 34:5084–5098, 2025a.

Zhoumin Lu, Yongbo Yu, Linru Ma, Feiping Nie, and Rong Wang. Capturing individuality and commonality between anchor graphs for multi-view clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 5860–5868, 8 2025b.

Chuan Tang, Miaomiao Li, Jun Wang, Chang Tang, En Zhu, and Xinwang Liu. Sample-level weighted and structure-enhanced anchor graph learning for scalable multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026.

Jun Wang, Miaomiao Li, Zhenglai Li, Hao Yu, Suyuan Liu, Dayu Hu, Chang Tang, and Xinwang Liu. Tensor multi-rank constraint guided anchor-wise adaptive alignment for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. *Advances in Neural Information Processing Systems*, 35:5882–5895, 2022a.

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, 2022b.

Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33: 4627–4639, 2024.

Yi Wen, Suyuan Liu, Xinhang Wan, Siwei Wang, Ke Liang, Xinwang Liu, Xihong Yang, and Pei Zhang. Efficient multi-view graph clustering with local and global structure preservation. In *Proceedings of the ACM International Conference on Multimedia*, pages 3021–3030, 2023.

Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):5187–5202, 2022.

Zengbiao Yang and Yihua Tan. The methods for improving large-scale multi-view clustering efficiency: A survey. *Artificial Intelligence Review*, 57:153, 2024.

Chao Zhang, Xiuyi Jia, Zechao Li, Chunlin Chen, and Huaxiong Li. Learning cluster-wise anchors for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38(15), pages 16696–16704, 2024.

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 AAAI Conference on Artificial Intelligence*, volume 37(9), pages 11262–11269, 2023.

Jiale Zou, Yan Chen, Bingbing Jiang, Peng Zhou, Liang Du, Lei Duan, and Yuhua Qian. Robust tensor learning with graph diffusion for scalable multi-view graph clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 2207–2215, 2025.

A Detailed derivation

A.1 Details of each distribution and the complete log-joint distribution

The Dirichlet prior is

$$p(\boldsymbol{\pi}) = \frac{\Gamma\Big(\sum_{k=1}^{K} \alpha_{0k}\Big)}{\prod_{k=1}^{K} \Gamma(\alpha_{0k})} \prod_{k=1}^{K} \pi_k^{\alpha_{0k}-1}, \qquad \boldsymbol{\pi} \in \Delta^{K-1}. \tag{26}$$

The categorical likelihood of the latent label is

$$p(z_i = k \mid \boldsymbol{\pi}) = \pi_k, \qquad k = 1, \ldots, K. \tag{27}$$

Equivalently, with one-hot indicators $z_{ik} = \mathbb{I}\{z_i = k\}$,

$$p(z_i \mid \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{ik}}. \tag{28}$$

We use the shape–rate parameterization for the Gamma distribution:

$$p(\tau_v) = \frac{b_0^{a_0}}{\Gamma(a_0)} \tau_v^{a_0-1} \exp(-b_0 \tau_v), \qquad \tau_v > 0. \tag{29}$$

Finally, conditional on $z_i = k$, the view-specific Gaussian observation density is

$$p(\boldsymbol{w}_i^{(v)} \mid z_i = k, \tau_v, \mathbf{C}^{(v)}) = \Big(\frac{\tau_v}{2\pi}\Big)^{m_v/2} \exp\Big[-\frac{\tau_v}{2} \Big\| \boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)} \Big\|_2^2\Big], \tag{30}$$

Equivalently, with one-hot indicators $z_{ik} = \mathbb{I}\{z_i = k\}$,

$$p(\boldsymbol{w}_i^{(v)} \mid z_i, \tau_v, \mathbf{C}^{(v)}) = \prod_{k=1}^{K} \left\{ \Big(\frac{\tau_v}{2\pi}\Big)^{m_v/2} \exp\Big[-\frac{\tau_v}{2} \Big\| \boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)} \Big\|_2^2\Big] \right\}^{z_{ik}}. \tag{31}$$

Introduce indicator variables $z_{ik} = \mathbf{1}[z_i = k]$, the complete log-joint can be written as

$$\begin{aligned} \log p(\mathcal{W}, z, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \{\mathbf{C}^{(v)}\}) = {} & \log \Gamma\left(\sum_{k=1}^{K} \alpha_{0k}\right) - \sum_{k=1}^{K} \log \Gamma(\alpha_{0k}) + \sum_{k=1}^{K} (\alpha_{0k} - 1) \log \pi_k \\ & + \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log \pi_k \\ & + \sum_{v=1}^{V} \Big(a_0 \log b_0 - \log \Gamma(a_0) + (a_0 - 1) \log \tau_v - b_0 \tau_v \Big) \\ & + \frac{1}{2} \sum_{v=1}^{V} \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \Big(m_v \log \tau_v - m_v \log(2\pi) - \tau_v D_{ik}^{(v)} \Big), \end{aligned} \tag{32}$$

where $D_{ik}^{(v)} = \|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}\|_2^2$.

A.2 Explicit ELBO expression

Taking expectation under the mean-field family and using $\mathbb{E}_q[z_{ik}] = \phi_{ik}$ gives

$$\begin{aligned}\mathcal{L} = \log\Gamma\left(\sum_{k=1}^{K}\alpha_{0k}\right) - \sum_{k=1}^{K}\log\Gamma(\alpha_{0k}) + \sum_{k=1}^{K}(\alpha_{0k}-1)\mathbb{E}_q[\log\pi_k] \\ + \sum_{i=1}^{n}\sum_{k=1}^{K}\phi_{ik}\mathbb{E}_q[\log\pi_k] \\ + \sum_{v=1}^{V}\Big(a_0\log b_0 - \log\Gamma(a_0) + (a_0-1)\mathbb{E}_q[\log\tau_v] - b_0\bar{\tau}_v\Big) \\ + \frac{1}{2}\sum_{v=1}^{V}\sum_{i=1}^{n}\sum_{k=1}^{K}\phi_{ik}\Big(m_v\mathbb{E}_q[\log\tau_v] - m_v\log(2\pi) - \bar{\tau}_v D_{ik}^{(v)}\Big) \\ + \mathcal{H}\big[q(\boldsymbol{\pi})\big] + \sum_{v=1}^{V}\mathcal{H}\big[q(\tau_v)\big] + \sum_{i=1}^{n}\mathcal{H}\big[q(z_i)\big]. \end{aligned} \tag{33}$$

For monitoring purposes, the entropies can be evaluated explicitly as

$$\begin{aligned}&\mathcal{H}\big[\mathrm{Dir}(\tilde{\boldsymbol{\alpha}})\big] = \log B(\tilde{\boldsymbol{\alpha}}) + (\sum_k \tilde{\alpha}_k - K)\psi(\sum_k \tilde{\alpha}_k) - \sum_{k=1}^{K}(\tilde{\alpha}_k - 1)\psi(\tilde{\alpha}_k), \\ &\mathcal{H}\big[\mathrm{Gamma}(\tilde{a}_v, \tilde{b}_v)\big] = \tilde{a}_v - \log\tilde{b}_v + \log\Gamma(\tilde{a}_v) + (1-\tilde{a}_v)\psi(\tilde{a}_v), \\ &\mathcal{H}\big[\mathrm{Cat}(\boldsymbol{\phi}_i)\big] = -\sum_{k=1}^{K}\phi_{ik}\log\phi_{ik}. \end{aligned} \tag{34}$$

Here $B(\tilde{\boldsymbol{\alpha}})$ is the multivariate beta function.

A.3 Derivation of the Dirichlet log-moment

The Dirichlet density on the probability simplex is

$$q_\pi(\boldsymbol{\pi}) = \frac{1}{B(\tilde{\boldsymbol{\alpha}})}\prod_{\ell=1}^{K}\pi_\ell^{\tilde{\alpha}_\ell - 1}, \tag{35}$$

where the multivariate beta function is

$$B(\tilde{\boldsymbol{\alpha}}) = \frac{\prod_{\ell=1}^{K}\Gamma(\tilde{\alpha}_\ell)}{\Gamma\Big(\sum_{\ell=1}^{K}\tilde{\alpha}_\ell\Big)}. \tag{36}$$

Equivalently, the normalization identity is

$$\int_{\Delta^{K-1}}\prod_{\ell=1}^{K}\pi_\ell^{\tilde{\alpha}_\ell - 1}\,d\boldsymbol{\pi} = B(\tilde{\boldsymbol{\alpha}}). \tag{37}$$

We differentiate both sides of the normalization identity with respect to $\tilde{\alpha}_k$. Under the usual regularity condition $\tilde{\alpha}_k > 0$, differentiation under the integral sign is valid. The derivative of the left-hand side is

$$\frac{\partial}{\partial\tilde{\alpha}_k}\int_{\Delta^{K-1}}\prod_{\ell=1}^{K}\pi_\ell^{\tilde{\alpha}_\ell - 1}\,d\boldsymbol{\pi} = \int_{\Delta^{K-1}}\log\pi_k\prod_{\ell=1}^{K}\pi_\ell^{\tilde{\alpha}_\ell - 1}\,d\boldsymbol{\pi}. \tag{38}$$

The derivative of the right-hand side is

$$\frac{\partial}{\partial\tilde{\alpha}_k}B(\tilde{\boldsymbol{\alpha}}). \tag{39}$$

Therefore,

$$\int_{\Delta^{K-1}} \log \pi_k \prod_{\ell=1}^{K} \pi_\ell^{\tilde{\alpha}_\ell - 1} \, d\boldsymbol{\pi} = \frac{\partial}{\partial \tilde{\alpha}_k} B(\tilde{\boldsymbol{\alpha}}). \tag{40}$$

Dividing both sides by $B(\tilde{\boldsymbol{\alpha}})$ gives

$$\frac{1}{B(\tilde{\boldsymbol{\alpha}})} \int_{\Delta^{K-1}} \log \pi_k \prod_{\ell=1}^{K} \pi_\ell^{\tilde{\alpha}_\ell - 1} \, d\boldsymbol{\pi} = \frac{1}{B(\tilde{\boldsymbol{\alpha}})} \frac{\partial}{\partial \tilde{\alpha}_k} B(\tilde{\boldsymbol{\alpha}}). \tag{41}$$

The left-hand side is exactly the expectation of $\log \pi_k$ under $q_\pi(\boldsymbol{\pi})$. Hence

$$\mathbb{E}_{q_\pi}[\log \pi_k] = \frac{\partial}{\partial \tilde{\alpha}_k} \log B(\tilde{\boldsymbol{\alpha}}). \tag{42}$$

It remains to compute this derivative explicitly. Since

$$\log B(\tilde{\boldsymbol{\alpha}}) = \sum_{\ell=1}^{K} \log \Gamma(\tilde{\alpha}_\ell) - \log \Gamma\left(\sum_{\ell=1}^{K} \tilde{\alpha}_\ell \right), \tag{43}$$

we have

$$\frac{\partial}{\partial \tilde{\alpha}_k} \log B(\tilde{\boldsymbol{\alpha}}) = \frac{\partial}{\partial \tilde{\alpha}_k} \log \Gamma(\tilde{\alpha}_k) - \frac{\partial}{\partial \tilde{\alpha}_k} \log \Gamma\left(\sum_{\ell=1}^{K} \tilde{\alpha}_\ell \right). \tag{44}$$

Using the definition of the digamma function,

$$\psi(x) = \frac{d}{dx} \log \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}, \tag{45}$$

we obtain

$$\frac{\partial}{\partial \tilde{\alpha}_k} \log B(\tilde{\boldsymbol{\alpha}}) = \psi(\tilde{\alpha}_k) - \psi\left(\sum_{\ell=1}^{K} \tilde{\alpha}_\ell \right). \tag{46}$$

Therefore,

$$\mathbb{E}_{q_\pi}[\log \pi_k] = \psi(\tilde{\alpha}_k) - \psi\left(\sum_{\ell=1}^{K} \tilde{\alpha}_\ell \right). \tag{47}$$

A.4 Derivation of the Gamma moments

Using the shape–rate parameterization, the Gamma density is

$$q_v(\tau_v) = \frac{\tilde{b}_v^{\tilde{a}_v}}{\Gamma(\tilde{a}_v)} \tau_v^{\tilde{a}_v - 1} \exp(-\tilde{b}_v \tau_v), \qquad \tau_v > 0, \tag{48}$$

with shape parameter $\tilde{a}_v > 0$ and rate parameter $\tilde{b}_v > 0$.

Derivation of $\mathbb{E}_q[\tau_v]$. By definition,

$$\mathbb{E}_q[\tau_v] = \int_0^\infty \tau_v q_v(\tau_v) \, d\tau_v. \tag{49}$$

Substituting the Gamma density gives

$$\mathbb{E}_q[\tau_v] = \int_0^\infty \tau_v \frac{\tilde{b}_v^{\tilde{a}_v}}{\Gamma(\tilde{a}_v)} \tau_v^{\tilde{a}_v - 1} \exp(-\tilde{b}_v \tau_v) \, d\tau_v = \frac{\tilde{b}_v^{\tilde{a}_v}}{\Gamma(\tilde{a}_v)} \int_0^\infty \tau_v^{\tilde{a}_v} \exp(-\tilde{b}_v \tau_v) \, d\tau_v. \tag{50}$$

Using the Gamma integral identity

$$\int_0^\infty x^{a-1} \exp(-bx) \, dx = \frac{\Gamma(a)}{b^a}, \qquad a > 0, \; b > 0, \tag{51}$$

with $a = \tilde{a}_v + 1$ and $b = \tilde{b}_v$, we obtain

$$\int_0^\infty \tau_v^{\tilde{a}_v} \exp(-\tilde{b}_v \tau_v)\, d\tau_v = \frac{\Gamma(\tilde{a}_v + 1)}{\tilde{b}_v^{\tilde{a}_v+1}}. \tag{52}$$

Therefore,

$$\mathbb{E}_q[\tau_v] = \frac{\tilde{b}_v^{\tilde{a}_v}}{\Gamma(\tilde{a}_v)} \frac{\Gamma(\tilde{a}_v + 1)}{\tilde{b}_v^{\tilde{a}_v+1}}. \tag{53}$$

Since $\Gamma(\tilde{a}_v + 1) = \tilde{a}_v \Gamma(\tilde{a}_v)$, we have

$$\mathbb{E}_q[\tau_v] = \frac{\tilde{a}_v}{\tilde{b}_v}. \tag{54}$$

Derivation of $\mathbb{E}_q[\log \tau_v]$. We next derive the logarithmic moment. Let

$$Z(\tilde{a}_v, \tilde{b}_v) = \int_0^\infty \tau_v^{\tilde{a}_v - 1} \exp(-\tilde{b}_v \tau_v)\, d\tau_v. \tag{55}$$

By the Gamma integral identity,

$$Z(\tilde{a}_v, \tilde{b}_v) = \frac{\Gamma(\tilde{a}_v)}{\tilde{b}_v^{\tilde{a}_v}}. \tag{56}$$

Differentiating $Z(\tilde{a}_v, \tilde{b}_v)$ with respect to $\tilde{a}_v$ gives

$$\frac{\partial}{\partial \tilde{a}_v} Z(\tilde{a}_v, \tilde{b}_v) = \int_0^\infty \log \tau_v\, \tau_v^{\tilde{a}_v - 1} \exp(-\tilde{b}_v \tau_v)\, d\tau_v. \tag{57}$$

Dividing both sides by $Z(\tilde{a}_v, \tilde{b}_v)$, we obtain

$$\mathbb{E}_q[\log \tau_v] = \frac{\partial}{\partial \tilde{a}_v} \log Z(\tilde{a}_v, \tilde{b}_v) = \frac{\partial}{\partial \tilde{a}_v} \left[\log \Gamma(\tilde{a}_v) - \tilde{a}_v \log \tilde{b}_v\right] = \psi(\tilde{a}_v) - \log \tilde{b}_v. \tag{58}$$

where ψ is the digamma function.

A.5 Derivation of the responsibility update

We solve this constrained maximization problem by introducing a Lagrange multiplier λ_i for the simplex constraint $\sum_{k=1}^K \phi_{ik} = 1$:

$$\mathcal{J}_i(\boldsymbol{\phi}_i, \lambda_i) = \sum_{k=1}^K \phi_{ik} \xi_{ik} - \sum_{k=1}^K \phi_{ik} \log \phi_{ik} + \lambda_i \left(\sum_{k=1}^K \phi_{ik} - 1 \right). \tag{59}$$

Taking derivative with respect to ϕ_{ik} and setting the gradient to zero yields

$$\frac{\partial \mathcal{J}_i}{\partial \phi_{ik}} = \xi_{ik} - (\log \phi_{ik} + 1) + \lambda_i = 0. \tag{60}$$

Equivalently,

$$\phi_{ik} = \exp(\xi_{ik} + \lambda_i - 1). \tag{61}$$

Using the simplex constraint,

$$1 = \sum_{k=1}^K \phi_{ik} = \exp(\lambda_i - 1) \sum_{k=1}^K \exp(\xi_{ik}). \tag{62}$$

Equivalently,

$$\exp(\lambda_i - 1) = \frac{1}{\sum_{\ell=1}^K \exp(\xi_{i\ell})}. \tag{63}$$

Therefore,

$$\phi_{ik} = \exp(\xi_{ik}) \exp(\lambda_i - 1) = \frac{\exp(\xi_{ik})}{\sum_{\ell=1}^K \exp(\xi_{i\ell})}. \tag{64}$$

A.6 Derivation of the prototype update

Fix view v and cluster k. The M-step subproblem is Eq. (24). Expanding the square gives

$$\sum_{i=1}^{n} \phi_{ik} \|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}\|_2^2 = \sum_{i=1}^{n} \phi_{ik} \left[(\boldsymbol{c}_k^{(v)})^\top \boldsymbol{c}_k^{(v)} - 2(\boldsymbol{c}_k^{(v)})^\top \boldsymbol{w}_i^{(v)} + (\boldsymbol{w}_i^{(v)})^\top \boldsymbol{w}_i^{(v)} \right]. \tag{65}$$

Differentiating with respect to $\boldsymbol{c}_k^{(v)}$ and setting the gradient to zero yields

$$2 \left(\sum_{i=1}^{n} \phi_{ik} \right) \boldsymbol{c}_k^{(v)} - 2 \sum_{i=1}^{n} \phi_{ik} \boldsymbol{w}_i^{(v)} = 0, \tag{66}$$

which gives Eq. (25). The Hessian is $2(\sum_i \phi_{ik})\mathbf{I}_{m_v}$, strictly positive definite because the denominator is positive. Thus the minimizer is unique.

A.7 Proof of convex hull and simplex closure

Proof of Proposition 4.1. Equation (25) can be rewritten as

$$\boldsymbol{c}_k^{(v)} = \sum_{i=1}^{n} \lambda_{ik}^{(v)} \boldsymbol{w}_i^{(v)}, \qquad \lambda_{ik}^{(v)} = \frac{\phi_{ik}}{\sum_{j=1}^{n} \phi_{jk}}. \tag{67}$$

Each coefficient $\lambda_{ik}^{(v)}$ is nonnegative and the coefficients sum to one. Hence $\boldsymbol{c}_k^{(v)}$ is a convex combination of the observed anchor vectors and therefore lies in their convex hull.

If each $\boldsymbol{w}_i^{(v)}$ lies in the simplex Δ^{m_v-1}, then every convex combination of them also lies in the simplex because the simplex is convex. Therefore $\boldsymbol{c}_k^{(v)} \in \Delta^{m_v-1}$. □

A.8 Proof of permutation invariance

Proof of Proposition 4.2. Let $\mathbf{P}$ be a permutation matrix, so $\mathbf{P}^\top \mathbf{P} = \mathbf{I}$. Then for every i and k,

$$\|\mathbf{P}\boldsymbol{w}_i^{(v)} - \mathbf{P}\boldsymbol{c}_k^{(v)}\|_2^2 = \|\mathbf{P}(\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)})\|_2^2 = (\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)})^\top \mathbf{P}^\top \mathbf{P} (\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}) = \|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}\|_2^2. \tag{68}$$

Therefore, the distances $D_{ik}^{(v)}$ are unchanged, and consequently, the ELBO, every variational update, and the final labels also remain unchanged.

For the prototype update,

$$\mathbf{P}\boldsymbol{c}_k^{(v)} = \mathbf{P} \left(\frac{\sum_i \phi_{ik} \boldsymbol{w}_i^{(v)}}{\sum_i \phi_{ik}} \right) = \frac{\sum_i \phi_{ik} \mathbf{P}\boldsymbol{w}_i^{(v)}}{\sum_i \phi_{ik}}, \tag{69}$$

which is exactly the prototype one would compute from the permuted coordinates. Therefore the entire trajectory is equivariant to any within-view anchor permutation, and the final labels are unchanged. □

A.9 Proof of adaptive weighting

Proof of Proposition 4.3. The posterior mean precision can be rewritten as

$$\bar{\tau}_v = \frac{a_0 + \frac{1}{2} n m_v}{b_0 + \frac{1}{2} SSE_v} = \left(\frac{b_0 + \frac{1}{2} SSE_v}{a_0 + \frac{1}{2} n m_v} \right)^{-1}. \tag{70}$$

Differentiate Eq. (70) with respect to SSE_v:

$$\frac{\partial \bar{\tau}_v}{\partial SSE_v} = -\frac{a_0 + \frac{1}{2} n m_v}{2\big(b_0 + \frac{1}{2} SSE_v\big)^2} < 0. \tag{71}$$

Thus $\bar{\tau}_v$ is strictly decreasing in SSE_v. Since $\bar{\tau}_v$ multiplies the view-specific distortion terms in Eq. (19), a lower residual sum of squares yields a larger contribution of that view to label assignment. □

B Full convergence proof

This appendix proves Theorem 4.4 in full detail. For notational convenience in the convergence proof, we use the block-wise concatenation $\mathbf{C} = [\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(V)}] \in \mathbb{R}^{K \times \sum_{v=1}^{V} m_v}$, where each block $\mathbf{C}^{(v)} \in \mathbb{R}^{K \times m_v}$ is the view-specific prototype matrix of view v. For two prototype collections $\mathbf{C}$ and $\mathbf{C}'$, the Frobenius norm is understood as $\|\mathbf{C} - \mathbf{C}'\|_F^2 = \sum_{v=1}^{V} \|\mathbf{C}^{(v)} - \mathbf{C}'^{(v)}\|_F^2$. This block-wise concatenation is only a notational device and does not impose or imply any cross-view anchor correspondence; each block remains in its own view-specific anchor coordinate system.

B.1 Compactness and distance bounds

For a fixed view v, let

$$\mathcal{W}_v = \{\boldsymbol{w}_1^{(v)}, \dots, \boldsymbol{w}_n^{(v)}\}, \qquad \Delta_v = \max_{1 \le i,j \le n} \|\boldsymbol{w}_i^{(v)} - \boldsymbol{w}_j^{(v)}\|_2. \tag{72}$$

Because the dataset is finite, $\Delta_v < \infty$.

Lemma B.1 (Prototype compactness and bounded distortions)**.** *Suppose the prototypes are initialized by Eq.* (25)*. Then for every iteration t, every view v, and every cluster k,*

$$\boldsymbol{c}_k^{(v,t)} \in \operatorname{conv}(\mathcal{W}_v), \qquad D_{ik}^{(v,t)} \le \Delta_v^2 \quad \textit{for all } i. \tag{73}$$

Proof. The first claim follows immediately from Proposition 4.1, since the initialization and every M-step update are convex combinations of observed anchor vectors. For the distance bound, write

$$\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v,t)} = \sum_{j=1}^{n} \lambda_{jk}^{(v,t)} (\boldsymbol{w}_i^{(v)} - \boldsymbol{w}_j^{(v)}), \tag{74}$$

where the $\lambda_{jk}^{(v,t)}$ are nonnegative and sum to one. Hence

$$\|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v,t)}\|_2 \le \sum_{j=1}^{n} \lambda_{jk}^{(v,t)} \|\boldsymbol{w}_i^{(v)} - \boldsymbol{w}_j^{(v)}\|_2 \le \Delta_v, \tag{75}$$

which gives $D_{ik}^{(v,t)} \le \Delta_v^2$. □

Lemma B.2 (Uniform bounds on Dirichlet and Gamma parameters)**.** *For every iteration t,*

$$\alpha_{0k} \le \tilde{\alpha}_k^{(t)} \le \alpha_{0k} + n, \tag{76}$$

$$\tilde{a}_v^{(t)} = a_0 + \frac{1}{2} n m_v, \tag{77}$$

$$b_0 \le \tilde{b}_v^{(t)} \le b_0 + \frac{1}{2} n \Delta_v^2, \tag{78}$$

$$\underline{\tau}_v \le \bar{\tau}_v^{(t)} \le \overline{\tau}_v, \qquad \underline{\tau}_v = \frac{a_0 + \frac{1}{2} n m_v}{b_0 + \frac{1}{2} n \Delta_v^2}, \quad \overline{\tau}_v = \frac{a_0 + \frac{1}{2} n m_v}{b_0}. \tag{79}$$

Proof. The bounds for $\tilde{\alpha}_k^{(t)}$ follow from Eq. (14) and the facts that $\phi_{ik}^{(t)} \ge 0$ and $\sum_k \phi_{ik}^{(t)} = 1$. Equation (77) is exactly Eq. (16). For $\tilde{b}_v^{(t)}$, Eq. (16) and Lemma B.1 give

$$SSE_v^{(t)} = \sum_{i=1}^{n} \sum_{k=1}^{K} \phi_{ik}^{(t)} D_{ik}^{(v,t)} \le \sum_{i=1}^{n} \left(\sum_{k=1}^{K} \phi_{ik}^{(t)} \right) \Delta_v^2 = n \Delta_v^2. \tag{80}$$

Substituting into Eq. (18) yields the stated bounds on $\bar{\tau}_v^{(t)}$. □

Lemma B.3 (Uniform interior bound on responsibilities)**.** *There exists $\varepsilon > 0$ such that for every iteration t and all i, k, $\phi_{ik}^{(t)} \ge \varepsilon$. Consequently, the cluster masses satisfy $N_k^{(t)} := \sum_{i=1}^{n} \phi_{ik}^{(t)} \ge n\varepsilon$ for all k, t.*

Proof. By Lemma B.2, all Dirichlet parameters $\tilde{\alpha}_k^{(t)}$ lie in compact intervals away from zero, and the total concentration $\sum_h \tilde{\alpha}_h^{(t)}$ lies in the compact interval $[A_0, A_0 + n]$, where $A_0 = \sum_h \alpha_{0h}$. Since the digamma function is continuous on $(0, \infty)$, there exists a finite constant B_π such that

$$\left| \psi(\tilde{\alpha}_k^{(t)}) - \psi\left(\sum_h \tilde{\alpha}_h^{(t)} \right) \right| \le B_\pi \qquad \text{for all } k, t. \tag{81}$$

Using Lemmas B.1 and B.2, the log-score in Eq. (19) satisfy

$$|\xi_{ik}^{(t)}| \le B_\pi + \frac{1}{2} \sum_{v=1}^{V} \overline{\tau}_v \Delta_v^2 =: B < \infty. \tag{82}$$

Therefore

$$\phi_{ik}^{(t)} = \frac{e^{\xi_{ik}^{(t)}}}{\sum_{h=1}^{K} e^{\xi_{ih}^{(t)}}} \ge \frac{e^{-B}}{K e^{B}} = \frac{e^{-2B}}{K} =: \varepsilon > 0. \tag{83}$$

Summing over i gives $N_k^{(t)} \ge n\varepsilon$. □

Lemmas B.1–B.3 imply that the state space

$$\mathcal{S} = \left\{ (\mathbf{\Phi}, \bar{\boldsymbol{\tau}}, \mathbf{C}) : \mathbf{\Phi} \in [\varepsilon, 1]^{n \times K},\ \mathbf{\Phi}\mathbf{1} = \mathbf{1},\ \underline{\tau}_v \le \bar{\tau}_v \le \overline{\tau}_v,\ \boldsymbol{c}_k^{(v)} \in \mathrm{conv}(\mathcal{W}_v) \right\} \tag{84}$$

is compact.

B.2 Monotonicity and exact improvement identities

Lemma B.4 (Monotonicity of one full iteration)**.** *Let $\mathcal{L}^{(t)} = \mathcal{L}(q^{(t)}, \mathbf{C}^{(t)})$ denote the ELBO after the t-th full outer iteration. Then $\mathcal{L}^{(t+1)} \ge \mathcal{L}^{(t)}$ for all t.*

Proof. Each block update is an exact optimizer of the ELBO over its own block with all other blocks fixed. Therefore every substep of the cyclic update is nondecreasing in ELBO, and so is their composition over one full iteration. Specifically,

$$\begin{aligned}
\mathcal{L}^{(t+1)} &= \mathcal{L}(q_\pi^{(t+1)}, q_z^{(t+1)}, q_\tau^{(t+1)}, \mathbf{C}^{(t+1)}) \\
&\ge \mathcal{L}(q_\pi^{(t+1)}, q_\tau^{(t+1)}, q_z^{(t+1)}, \mathbf{C}^{(t)}) \\
&\ge \mathcal{L}(q_\pi^{(t+1)}, q_\tau^{(t+1)}, q_z^{(t)}, \mathbf{C}^{(t)}) \\
&\ge \mathcal{L}(q_\pi^{(t+1)}, q_\tau^{(t)}, q_z^{(t)}, \mathbf{C}^{(t)}) \\
&\ge \mathcal{L}(q_\pi^{(t)}, q_\tau^{(t)}, q_z^{(t)}, \mathbf{C}^{(t)}) = \mathcal{L}^{(t)}.
\end{aligned} \tag{85}$$

□

Lemma B.5 (Exact gain of the responsibility block)**.** *Let $\Delta_\Phi^{(t)}$ denote the ELBO improvement contributed by the responsibility update at iteration $t+1$. Then*

$$\Delta_\Phi^{(t)} = \sum_{i=1}^{n} \mathrm{KL}\big(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)}\big) \ge 0. \tag{86}$$

Equality holds if and only if $\boldsymbol{\phi}_i^{(t)} = \boldsymbol{\phi}_i^{(t+1)}$ for all i.

Proof. We derive the exact ELBO gain obtained by updating the responsibility block. Consider the update from $\Phi^{(t)}$ to $\Phi^{(t+1)}$, while all other blocks used to form the responsibility scores are fixed. Define

$$\xi_{ik}^{(t+1)} = \mathbb{E}_{q_\pi^{(t+1)}}[\log \pi_k] - \frac{1}{2} \sum_{v=1}^{V} \bar{\tau}_v^{(t+1)} \left\| \boldsymbol{w}_i^{(v)} - \mathbf{c}_k^{(v,t)} \right\|_2^2, \tag{87}$$

Here $\xi_{ik}^{(t+1)}$ denotes the fixed score used in the responsibility update at iteration $t+1$. The superscript only records the iteration at which this score is evaluated.

With all other blocks fixed, the part of the ELBO depending on Φ is

$$\mathcal{L}_z(\mathbf{\Phi}; \mathbf{\Xi}^{(t+1)}) = \sum_{i=1}^{n} \left[\sum_{k=1}^{K} \phi_{ik} \xi_{ik}^{(t+1)} - \sum_{k=1}^{K} \phi_{ik} \log \phi_{ik} \right] + \text{const}, \tag{88}$$

where the constant is independent of $\mathbf{\Phi}$. The responsibility update is

$$\phi_{ik}^{(t+1)} = \frac{\exp(\xi_{ik}^{(t+1)})}{\sum_{\ell=1}^{K} \exp(\xi_{i\ell}^{(t+1)})}. \tag{89}$$

Define the local log-partition function

$$A_i^{(t+1)} = \log \sum_{\ell=1}^{K} \exp(\xi_{i\ell}^{(t+1)}). \tag{90}$$

Then

$$\log \phi_{ik}^{(t+1)} = \xi_{ik}^{(t+1)} - A_i^{(t+1)}, \qquad \xi_{ik}^{(t+1)} = \log \phi_{ik}^{(t+1)} + A_i^{(t+1)} \tag{91}$$

For each sample i, define the local objective

$$\ell_i(\boldsymbol{\phi}_i; \xi_i^{(t+1)}) = \sum_{k=1}^{K} \phi_{ik} \xi_{ik}^{(t+1)} - \sum_{k=1}^{K} \phi_{ik} \log \phi_{ik}. \tag{92}$$

Evaluating it at the updated responsibility vector gives

$$\begin{aligned}
\ell_i(\boldsymbol{\phi}_i^{(t+1)}; \xi_i^{(t+1)}) &= \sum_{k=1}^{K} \phi_{ik}^{(t+1)} \xi_{ik}^{(t+1)} - \sum_{k=1}^{K} \phi_{ik}^{(t+1)} \log \phi_{ik}^{(t+1)} \\
&= \sum_{k=1}^{K} \phi_{ik}^{(t+1)} \xi_{ik}^{(t+1)} - \sum_{k=1}^{K} \phi_{ik}^{(t+1)} \left(\xi_{ik}^{(t+1)} - A_i^{(t+1)} \right) \\
&= A_i^{(t+1)} \sum_{k=1}^{K} \phi_{ik}^{(t+1)} \\
&= A_i^{(t+1)}.
\end{aligned} \tag{93}$$

Next, evaluating the same local objective at the old responsibility vector, we obtain

$$\begin{aligned}
\ell_i(\boldsymbol{\phi}_i^{(t)}; \xi_i^{(t+1)}) &= \sum_{k=1}^{K} \phi_{ik}^{(t)} \xi_{ik}^{(t+1)} - \sum_{k=1}^{K} \phi_{ik}^{(t)} \log \phi_{ik}^{(t)} \\
&= \sum_{k=1}^{K} \phi_{ik}^{(t)} \left(\log \phi_{ik}^{(t+1)} + A_i^{(t+1)} \right) - \sum_{k=1}^{K} \phi_{ik}^{(t)} \log \phi_{ik}^{(t)} \\
&= A_i^{(t+1)} \sum_{k=1}^{K} \phi_{ik}^{(t)} + \sum_{k=1}^{K} \phi_{ik}^{(t)} \log \phi_{ik}^{(t+1)} - \sum_{k=1}^{K} \phi_{ik}^{(t)} \log \phi_{ik}^{(t)} \\
&= A_i^{(t+1)} - \sum_{k=1}^{K} \phi_{ik}^{(t)} \log \frac{\phi_{ik}^{(t)}}{\phi_{ik}^{(t+1)}} \\
&= A_i^{(t+1)} - \mathrm{KL}\left(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)} \right).
\end{aligned} \tag{94}$$

Subtracting the two expressions gives

$$\ell_i(\boldsymbol{\phi}_i^{(t+1)}; \xi_i^{(t+1)}) - \ell_i(\boldsymbol{\phi}_i^{(t)}; \xi_i^{(t+1)}) = \mathrm{KL}\left(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)} \right). \tag{95}$$

Summing over $i = 1, \ldots, n$, we obtain

$$\begin{aligned}\Delta_{\Phi}^{(t)} &:= \mathcal{L}_z(\Phi^{(t+1)}; \xi^{(t+1)}) - \mathcal{L}_z(\Phi^{(t)}; \xi^{(t+1)}) \\ &= \sum_{i=1}^{n} \left[\ell_i(\boldsymbol{\phi}_i^{(t+1)}; \xi_i^{(t+1)}) - \ell_i(\boldsymbol{\phi}_i^{(t)}; \xi_i^{(t+1)}) \right] \\ &= \sum_{i=1}^{n} \mathrm{KL}\left(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)} \right) \\ &\geq 0. \end{aligned} \tag{96}$$

The nonnegativity follows from the properties of KL divergence. Equality holds if and only if $\mathrm{KL}\left(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)} \right) = 0$ for every i, which is equivalent to $\boldsymbol{\phi}_i^{(t)} = \boldsymbol{\phi}_i^{(t+1)}$ for all i. □

Lemma B.6 (Exact gain of the prototype block)**.** *Let $\Delta_{\mathbf{C}}^{(t)}$ denote the ELBO improvement contributed by the prototype update at iteration t. Then*

$$\Delta_{\mathbf{C}}^{(t)} = \frac{1}{2} \sum_{v=1}^{V} \bar{\tau}_v^{(t+1)} \sum_{k=1}^{K} N_k^{(t+1)} \left\| \boldsymbol{c}_k^{(v,t+1)} - \boldsymbol{c}_k^{(v,t)} \right\|_2^2 \geq \frac{1}{2} \underline{\tau}\, n\varepsilon \left\| \mathbf{C}^{(t+1)} - \mathbf{C}^{(t)} \right\|_F^2, \quad \underline{\tau} := \min_{1 \leq v \leq V} \tau_v. \tag{97}$$

Proof. With $\Phi^{(t+1)}$ and $\bar{\boldsymbol{\tau}}^{(t+1)}$ fixed, the prototype-dependent part of the negative ELBO is

$$F(\mathbf{C}) = \frac{1}{2} \sum_{v=1}^{V} \bar{\tau}_v^{(t+1)} \sum_{i=1}^{n} \sum_{k=1}^{K} \phi_{ik}^{(t+1)} \left\| \boldsymbol{w}_i^{(v)} - \mathbf{c}_k^{(v)} \right\|_2^2. \tag{98}$$

Therefore, increasing the ELBO with respect to $\mathbf{C}$ is equivalent to decreasing $F(\mathbf{C})$.

For each view v and cluster k, the subproblem is

$$\min_{\mathbf{c}_k^{(v)} \in \mathbb{R}^{m_v}} F_{vk}(\mathbf{c}_k^{(v)}) := \frac{1}{2} \bar{\tau}_v^{(t+1)} \sum_{i=1}^{n} \phi_{ik}^{(t+1)} \left\| \boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)} \right\|_2^2. \tag{99}$$

Define the effective cluster mass

$$N_k^{(t+1)} = \sum_{i=1}^{n} \phi_{ik}^{(t+1)}. \tag{100}$$

Because the softmax responsibilities are strictly positive, $N_k^{(t+1)} > 0$. The closed-form minimizer is

$$\boldsymbol{c}_k^{(v,t+1)} = \frac{\sum_{i=1}^{n} \phi_{ik}^{(t+1)} \boldsymbol{w}_i^{(v)}}{N_k^{(t+1)}}. \tag{101}$$

Fix a view v and a cluster k. For readability, temporarily define

$$a_i = \phi_{ik}^{(t+1)}, \qquad \boldsymbol{x}_i = \boldsymbol{w}_i^{(v)}, \qquad \boldsymbol{m} = \boldsymbol{c}_k^{(v,t+1)} = \frac{\sum_{i=1}^{n} a_i \boldsymbol{x}_i}{\sum_{i=1}^{n} a_i}. \tag{102}$$

Here $\boldsymbol{m}$ is only a temporary shorthand for the updated prototype $\boldsymbol{c}_k^{(v,t+1)}$. Let

$$A = \sum_{i=1}^{n} a_i = N_k^{(t+1)}. \tag{103}$$

For any vector $\boldsymbol{c} \in \mathbb{R}^{m_v}$, using $\boldsymbol{x}_i - \boldsymbol{c} = (\boldsymbol{x}_i - \boldsymbol{m}) + (\boldsymbol{m} - \boldsymbol{c})$, we have

$$\sum_{i=1}^{n} a_i \|\boldsymbol{x}_i - \boldsymbol{c}\|_2^2 = \sum_{i=1}^{n} a_i \|\boldsymbol{x}_i - \boldsymbol{m}\|_2^2 + A\|\boldsymbol{m} - \boldsymbol{c}\|_2^2 + 2 \left\langle \sum_{i=1}^{n} a_i (\boldsymbol{x}_i - \boldsymbol{m}), \boldsymbol{m} - \boldsymbol{c} \right\rangle. \tag{104}$$

The cross term is zero because

$$\sum_{i=1}^{n} a_i(\boldsymbol{x}_i - \boldsymbol{m}) = \sum_{i=1}^{n} a_i \boldsymbol{x}_i - A\boldsymbol{m} = \mathbf{0}. \tag{105}$$

Therefore,

$$\sum_{i=1}^{n} a_i \|\boldsymbol{x}_i - \boldsymbol{c}\|_2^2 - \sum_{i=1}^{n} a_i \|\boldsymbol{x}_i - \boldsymbol{m}\|_2^2 = A\|\boldsymbol{c} - \boldsymbol{m}\|_2^2. \tag{106}$$

Now substitute back

$$\boldsymbol{c} = \boldsymbol{c}_k^{(v,t)}, \qquad \boldsymbol{m} = \boldsymbol{c}_k^{(v,t+1)}, \qquad A = N_k^{(t+1)}. \tag{107}$$

For the (v, k)-th prototype subproblem, we obtain

$$F_{vk}(\boldsymbol{c}_k^{(v,t)}) - F_{vk}(\boldsymbol{c}_k^{(v,t+1)}) = \frac{1}{2}\bar{\tau}_v^{(t+1)} N_k^{(t+1)} \left\| \boldsymbol{c}_k^{(v,t)} - \boldsymbol{c}_k^{(v,t+1)} \right\|_2^2. \tag{108}$$

Summing this identity over all views $v = 1, \ldots, V$ and clusters $k = 1, \ldots, K$ gives

$$F(\mathbf{C}^{(t)}) - F(\mathbf{C}^{(t+1)}) = \frac{1}{2} \sum_{v=1}^{V} \bar{\tau}_v^{(t+1)} \sum_{k=1}^{K} N_k^{(t+1)} \left\| \boldsymbol{c}_k^{(v,t)} - \boldsymbol{c}_k^{(v,t+1)} \right\|_2^2. \tag{109}$$

Since the ELBO differs from $-F(\mathbf{C})$ only by terms independent of $\mathbf{C}$, we have

$$\Delta_{\mathbf{C}}^{(t)} = \mathcal{L}(\mathbf{C}^{(t+1)}) - \mathcal{L}(\mathbf{C}^{(t)}) = F(\mathbf{C}^{(t)}) - F(\mathbf{C}^{(t+1)}). \tag{110}$$

The lower bound in Eq. (97) follows from Lemma B.3, which gives $N_k^{(t+1)} \geq n\varepsilon$, and from Lemma B.2, which gives $\bar{\tau}_v^{(t+1)} \geq \underline{\tau}_v \geq \underline{\tau}$. □

B.3 Proof of the convergence theorem

Proof of Theorem 4.4. Building on the previous analysis, we further prove part(i)-(iv).

Part (i): The ELBO sequence is nondecreasing and convergent. By Lemmas B.1–B.3, every iterate lies in the compact state space $\mathcal{S}$. The ELBO is a continuous function of $(\boldsymbol{\Phi}, \bar{\boldsymbol{\tau}}, \mathbf{C})$ on $\mathcal{S}$ because all parameters stay in compact subsets away from singularities. Therefore the ELBO is bounded above on $\mathcal{S}$. By Lemma B.4, the sequence $\{\mathcal{L}^{(t)}\}$ is nondecreasing and bounded above, hence convergent. This proves part (i).

Part (ii): Every prototype update is well defined without any exact empty cluster. Part (ii) is exactly Lemma B.3.

Part (iii): The successive differences vanish. To prove part (iii), note that

$$0 \leq \Delta_{\Phi}^{(t)} \leq \mathcal{L}^{(t+1)} - \mathcal{L}^{(t)} \to 0, \tag{111}$$

because the total ELBO increase over one iteration tends to zero when the ELBO sequence converges. By Lemma B.5,

$$\sum_{i=1}^{n} \mathrm{KL}\big(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)}\big) \to 0. \tag{112}$$

Pinsker's inequality yields, for every i,

$$\|\boldsymbol{\phi}_i^{(t+1)} - \boldsymbol{\phi}_i^{(t)}\|_1^2 \leq 2\mathrm{KL}\big(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)}\big). \tag{113}$$

Summing over i gives

$$\|\boldsymbol{\Phi}^{(t+1)} - \boldsymbol{\Phi}^{(t)}\|_F^2 \leq \sum_{i=1}^{n} \|\boldsymbol{\phi}_i^{(t+1)} - \boldsymbol{\phi}_i^{(t)}\|_1^2 \leq 2\sum_{i=1}^{n} \mathrm{KL}\big(\boldsymbol{\phi}_i^{(t)} \,\|\, \boldsymbol{\phi}_i^{(t+1)}\big) \to 0. \tag{114}$$

Thus $\|\boldsymbol{\Phi}^{(t+1)} - \boldsymbol{\Phi}^{(t)}\|_F \to 0$.

Similarly,

$$0 \le \Delta_{\mathbf{C}}^{(t)} \le \mathcal{L}^{(t+1)} - \mathcal{L}^{(t)} \to 0. \tag{115}$$

Applying Lemma B.6 yields

$$\|\mathbf{C}^{(t+1)} - \mathbf{C}^{(t)}\|_F^2 \le \frac{2}{\underline{\tau}\, n\varepsilon}\, \Delta_{\mathbf{C}}^{(t)} \to 0. \tag{116}$$

Hence $\|\mathbf{C}^{(t+1)} - \mathbf{C}^{(t)}\|_F \to 0$.

It remains to show that $\|\bar{\boldsymbol{\tau}}^{(t+1)} - \bar{\boldsymbol{\tau}}^{(t)}\|_2 \to 0$. Define the continuous map

$$H_v(\boldsymbol{\Phi}, \mathbf{C}) = \frac{a_0 + \frac{1}{2} n m_v}{b_0 + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{K} \phi_{ik} \|\boldsymbol{w}_i^{(v)} - \boldsymbol{c}_k^{(v)}\|_2^2}. \tag{117}$$

Then $\bar{\tau}_v^{(t+1)} = H_v(\boldsymbol{\Phi}^{(t)}, \mathbf{C}^{(t)})$. Because $\mathcal{S}$ is compact, $H = (H_1, \ldots, H_V)$ is uniformly continuous. We have already shown that

$$\|\boldsymbol{\Phi}^{(t)} - \boldsymbol{\Phi}^{(t-1)}\|_F \to 0, \qquad \|\mathbf{C}^{(t)} - \mathbf{C}^{(t-1)}\|_F \to 0. \tag{118}$$

Therefore

$$\|(\boldsymbol{\Phi}^{(t)}, \mathbf{C}^{(t)}) - (\boldsymbol{\Phi}^{(t-1)}, \mathbf{C}^{(t-1)})\| \to 0, \tag{119}$$

which implies by uniform continuity that

$$\|\bar{\boldsymbol{\tau}}^{(t+1)} - \bar{\boldsymbol{\tau}}^{(t)}\|_2 = \|H(\boldsymbol{\Phi}^{(t)}, \mathbf{C}^{(t)}) - H(\boldsymbol{\Phi}^{(t-1)}, \mathbf{C}^{(t-1)})\|_2 \to 0. \tag{120}$$

This proves part (iii).

Part (iv): Every accumulation point is a fixed point. Let $\mathcal{S}^{(t)} = (\boldsymbol{\Phi}^{(t)}, \bar{\boldsymbol{\tau}}^{(t)}, \mathbf{C}^{(t)})$ denote the reduced state at iteration t, and let $\mathcal{T}$ denote one full iteration of our algorithm. Thus $\mathcal{S}^{(t+1)} = \mathcal{T}(\mathcal{S}^{(t)})$.

Let $\mathcal{S}^\star$ be an accumulation point of the sequence $\{\mathcal{S}^{(t)}\}_{t \ge 0}$. By definition of an accumulation point, there exists a strictly increasing sequence of iteration indices $0 \le t_1 < t_2 < \cdots$, such that $\mathcal{S}^{(t_j)} \to \mathcal{S}^\star$ as $j \to \infty$. Here j indexes the subsequence, while t_j is the actual algorithmic iteration number selected by that subsequence. Since $t_j \to \infty$ as $j \to \infty$, any asymptotic property that holds as $t \to \infty$ also holds along the subsequence $t = t_j$.

From the previous part of the proof, we have

$$\|\mathcal{S}^{(t+1)} - \mathcal{S}^{(t)}\| \to 0 \qquad \text{as } t \to \infty. \tag{121}$$

Therefore, along the subsequence,

$$\|\mathcal{S}^{(t_j+1)} - \mathcal{S}^{(t_j)}\| \to 0 \qquad \text{as } j \to \infty. \tag{122}$$

Since $\mathcal{S}^{(t_j)} \to \mathcal{S}^\star$, it follows that

$$\mathcal{S}^{(t_j+1)} = \mathcal{S}^{(t_j)} + \big(\mathcal{S}^{(t_j+1)} - \mathcal{S}^{(t_j)}\big) \to \mathcal{S}^\star. \tag{123}$$

On the other hand, by the definition of the full update map $\mathcal{T}$,

$$\mathcal{S}^{(t_j+1)} = \mathcal{T}(\mathcal{S}^{(t_j)}). \tag{124}$$

The update map $\mathcal{T}$ is continuous on the considered compact set: the Dirichlet update is a finite sum of responsibilities, the Gamma update is a continuous function of the weighted residual sums of squares, the responsibility update is a softmax of finite scores, and the prototype update is a weighted average with strictly positive denominators. Hence, taking the limit $j \to \infty$ gives

$$\lim_{j \to \infty} \mathcal{S}^{(t_j+1)} = \lim_{j \to \infty} \mathcal{T}(\mathcal{S}^{(t_j)}) = \mathcal{T}(\mathcal{S}^\star). \tag{125}$$

Since $\mathcal{S}^{(t_j+1)} \to \mathcal{S}^\star$ has already been shown, we know that

$$\mathcal{T}(\mathcal{S}^\star) = \mathcal{S}^\star. \tag{126}$$

Thus every accumulation point of the sequence is a fixed point of one full iteration of our algorithm.

It remains to justify why such a fixed point of the full cyclic update map is also coordinatewise stationary. Starting from $\mathcal{S}^\star$, one full cycle of our algorithm returns exactly to $\mathcal{S}^\star$. Hence the ELBO value at the beginning and at the end of the cycle is the same. On the other hand, each substep of the cycle is an exact coordinate maximization and therefore is nondecreasing in the ELBO. Since the total ELBO increase over the full cycle is zero, every individual substep must also have zero ELBO increase. The coordinate maximizers are unique on the compact state space considered above: the Dirichlet and Gamma variational factors are uniquely determined by their updated parameters, the responsibility block has a unique softmax solution because of the entropy term, and the prototype block has a unique weighted least-squares solution because the effective cluster masses are uniformly bounded away from zero. Therefore, zero increase in each substep implies that each coordinate block already maximizes the ELBO with respect to that block, with all other blocks held fixed. Consequently, $\mathcal{S}^\star$ satisfies the coordinate-optimality conditions for the responsibility block, the Dirichlet factor, the Gamma factors, and the prototype block. Thus $\mathcal{S}^\star$ is a coordinatewise stationary point of the variational EM objective. □

C Algorithm overview

For algorithmic implementation, we write the view-specific anchor graph as the row-stacked matrix $\boldsymbol{W}^{(v)} = [\boldsymbol{w}_1^{(v)}, \ldots, \boldsymbol{w}_n^{(v)}]^\top \in \mathbb{R}^{n \times m_v}$, whose i-th row is the sample-to-anchor vector $\boldsymbol{w}_i^{(v)}$. Thus our algorithm takes the collection $\{\boldsymbol{W}^{(v)}\}_{v=1}^V$, equivalently $\mathcal{W}$, as input. Unless prior knowledge is available, we use a noninformative symmetric setting $\boldsymbol{\alpha}_0 = \mathbf{1}_K$ for the Dirichlet prior and set $a_0 = b_0 = 1$ for the Gamma precision prior. In practice, these hyperparameters usually have limited influence after the posterior updates, especially when the number of samples and anchor dimensions are moderate to large, because the updates are dominated by the data-dependent terms $\sum_i \phi_{ik}$, $nm_v/2$, and $\mathrm{SSE}_v/2$. The specific algorithm flow is described in Algorithm 1.

Algorithm 1 Probabilistic Multi-View Anchor Graph Clustering (PMVAGC)

Require: Multi-view anchor graphs $\{\mathbf{W}^{(v)}\}_{v=1}^V$, cluster number K.

1: Initialize responsibility matrix $\boldsymbol{\Phi}$ and prototypes $\{\mathbf{C}^{(v)}\}$
2: **repeat**
3: Update $q(\boldsymbol{\pi})$ by Eq. (14), and compute $\mathbb{E}_q[\log \pi_k]$ by Eq. (17)
4: Update $q(\tau_v)$ by Eq. (16), and compute $\bar{\tau}_v = \mathbb{E}_q[\tau_v]$ and $\mathbb{E}_q[\log \tau_v]$ by Eq. (18)
5: Update $\boldsymbol{\phi}_i$ using Eq. (22), with the reduced log-score $\boldsymbol{\xi}_i$ in Eq. (19)
6: Update $\boldsymbol{c}_k^{(v)}$ by Eq. (25)
7: **until** relative ELBO increase or parameter change falls below a tolerance
8: **return** soft assignments $\boldsymbol{\Phi}$, prototypes $\{\mathbf{C}^{(v)}\}$, and hard labels $\hat{z}_i = \arg\max_k \phi_{ik}$

D Experimental details

D.1 Datasets

MSRC is an image dataset with multiple visual categories. ORL consists of face images captured at different times under varying illumination and facial expressions. The images show open or closed eyes, smiling or non-smiling faces, and subjects with or without glasses. 100Leaves contains plant leaf images, with each category corresponding to a plant species and containing 16 leaf samples. Its three views comprise shape descriptors, fine-scale margin features, and texture histograms. NGs is a subset of 20 Newsgroups containing discussion posts grouped by topic. MNIST contains images of handwritten digits from 0 to 9. YTF10, YTF20, and YTF50 contain face frames from YouTube Faces, with local binary patterns, histograms of oriented gradients, GIST, and Gabor features as views. CIFAR10 and CIFAR100 contain small color images with an original resolution of 32 by 32 pixels and are each represented by three feature views extracted using three different pretrained neural networks. CIFAR10 includes categories such as airplanes, cats, and trucks; CIFAR100 also covers flowers, household furniture, and outdoor scenes. Table 4 lists dataset statistics.

Table 4: Dataset descriptions

Dataset	#View	#Instance	#Class	#Feature
MSRC	6	210	7	1302, 48, 512, 100, 256, 210
ORL	3	400	40	4096, 3304, 6750
100Leaves	3	1600	100	64, 64, 64
NGs	3	500	5	2000, 2000, 2000
MNIST	3	60000	10	342, 1024, 64
YTF10	4	38654	10	944, 576, 512, 640
YTF20	4	63896	20	944, 576, 512, 640
YTF50	4	126054	50	944, 576, 512, 640
CIFAR10	3	50000	10	512, 2048, 1024
CIFAR100	3	50000	100	512, 2048, 1024

D.2 Compared methods

We briefly summarize the compared methods. These methods cover representative strategies for scalable multi-view anchor graph clustering, including shared-anchor construction, independently learned view-specific anchor graphs, explicit anchor alignment, weighted graph fusion, and tensor-regularized cross-view coupling.

- **LMVSC** independently learns one anchor graph for each view through anchor-based subspace clustering, and then equally concatenates the resulting view-specific anchor graphs for the final clustering.
- **FPMVS** performs multi-view subspace learning with a shared anchor graph across all views. In this way, all views are coordinated through the same anchor graph representation.
- **EMVGC** also learns a shared anchor graph across all views for subspace learning, and further imposes local coupling between view-specific anchor structures and the shared anchor graph.
- **SFMC** first concatenates features from all views and selects samples with unified indices as anchors through a feature-scoring strategy. The corresponding anchor graphs are then constructed for different views and fused in a weighted manner.
- **FDAGF** separately learns multiple anchor-based subspace representations for each view. Within each view, multiple tasks may use diverse anchor sizes, and all learned anchor graphs are finally concatenated for clustering.
- **FMVACC** first performs anchor-based subspace learning independently for each view, and then conducts anchor graph alignment and fusion. Its key design lies in the formulation of the alignment objective.
- **OMSC** performs subspace learning with a shared anchor graph across all views, and further imposes cluster-level regularization on the learned anchor graph.
- **AEVC** first learns an anchor-based subspace representation independently for each view, then aligns both anchors and anchor graphs across views, and finally learns a shared consensus anchor graph.
- **ESTMC** jointly learns view-specific anchor graphs, where each view has its own anchor graph. The anchor graphs are directly coupled through tensor-norm regularization, thereby inducing implicit cross-view alignment.
- **MVS4C** concatenates features from all views and selects samples with unified indices as anchors to construct view-specific anchor graphs. These graphs are then fused in a one-step weighted manner.
- **RTGD** constructs an individual anchor graph for each view and organizes the resulting structures as bipartite graphs. Cross-view information is then propagated through cross-view diffusion and tensor-level learning.
- **TMC** jointly learns view-specific anchor graphs, with each view maintaining its own anchor graph. Both anchors and anchor graphs are directly coupled by tensor-norm regularization to encourage implicit alignment.
- **SLWSE** performs subspace learning with a shared anchor graph across all views, and further imposes coupling related to sample-cluster assignments on the learned anchor graph.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the main claims, including the alignment-free probabilistic formulation, view-specific anchor spaces and anchor numbers, closed-form variational EM updates, convergence guarantees, and empirical effectiveness and efficiency. These claims are supported by the theoretical analysis and experimental results.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses the input anchor graphs used by PMVAGC and provides a complexity analysis showing the dependence on the number of clusters and total anchor cardinality. Like other anchor-based clustering methods, PMVAGC may be affected by input anchor graph quality and by the choice of anchor numbers.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The assumptions are stated in the theorem and propositions, and complete derivations and proofs are provided in the appendices, including the variational updates, prototype properties, adaptive weighting, and convergence analysis.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper reports the datasets, compared methods, evaluation metrics, anchor-number search range, number of runs, hardware configuration, and algorithmic procedure needed to reproduce the reported experimental results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide MATLAB code and a one-click reproducible demo on the MSRC dataset in the supplementary material. All datasets are publicly available.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental setup specifies the datasets, compared methods, evaluation metrics, anchor graph construction, anchor-number search range, number of runs, and hardware configuration.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The main performance table reports the mean performance and standard deviation over five runs for the compared methods.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper reports the hardware used for all experiments, including the CPU model and memory size, and further provides runtime comparisons in the main experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses publicly available benchmark datasets and standard clustering evaluation protocols, and does not involve human-subject experiments, private data, or safety-critical deployment.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This work focuses on a general-purpose clustering algorithm and does not introduce a deployed system or application-specific decision pipeline with direct societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release high-risk models, scraped datasets, or generative systems that require special safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [No]

Justification: The paper cites the related methods and uses publicly available datasets, but the current version does not explicitly list the licenses or terms of use for all existing datasets.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We provide MATLAB code in the supplementary material, including a one-click MSRC demo script.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or human-subject studies.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve human-subject research and therefore does not require IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development, theoretical analysis, and experiments do not involve LLMs as an important, original, or non-standard component.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.