Pseudo Multi-view K-means Clustering

Jinqian Chen[1], **Jihua Zhu**[1*], **Haoyu Tang**[2], **Qinghai Zheng**[3]

[1]School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]School of Software, Shandong University, Jinan 250101, China.
[3]College of Computer and Data Science, Fuzhou University, Fuzhou 350108, China
chenjinqian@stu.xjtu.edu.cn, zhujh@xjtu.edu.cn, tanghao258@sdu.edu.cn, zhengqinghai@fzu.edu.cn

Abstract

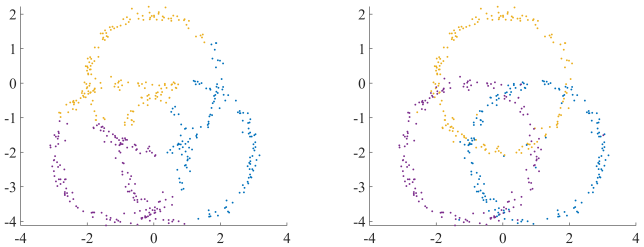
Clustering with k-means is well-established and efficient, but often struggles with complex data distributions because the clustering performance hinges on how well the centroids capture the data distribution, and conventional k-means usually fails to produce representative centroids under such conditions. To address this limitation, we propose **P**seudo **M**ulti-view **K**-means **C**lustering (PMKC), a novel framework that simulates a multi-view learning paradigm within a single-view setting by generating multiple soft k-means decompositions. Each decomposition can be treated as an individual view and investigates a distinct perspective of the data. Specifically, to encourage complementary structure, we impose an independence constraint among cluster centers, and to integrate these diverse clusterings, we model the soft assignment matrices as a third-order tensor and apply low-rank regularization to extract a shared latent structure. This design not only enhances clustering robustness but also improves the stability and consistency of the final results. Experimental results on several benchmark datasets demonstrate that PMKC achieves superior clustering performance compared to state-of-the-art methods.

Introduction

As a core component of unsupervised learning, clustering is widely used to group data samples into meaningful clusters based on their inherent structure and usually serves as a preprocessing step for subsequent machine learning tasks (Nie, Wang, and Li 2019; Oyewole and Thopil 2023). Over the years, numerous clustering algorithms have been proposed, ranging from prototype-based methods, *e.g.*, k-means (Ahmed, Seraj, and Islam 2020), graph-based methods, *e.g.*, spectral clustering (Von Luxburg 2007), to subspace and matrix factorization-based models (Elhamifar and Vidal 2013; Liu et al. 2012; Li et al. 2025). Among various clustering algorithms, k-means remains one of the most widely adopted methods due to its conceptual simplicity and computational efficiency (Arthur and Vassilvitskii 2006; Yin et al. 2024; Ren, Sun, and Wei 2021).

However, despite its popularity, k-means often exhibits poor performance when applied to data with complex distributions, such as non-convex structures, overlapping clusters, or heterogeneous densities. Taking a three circles dataset in Fig. 1 as an example, we can observe that k-means has difficulty achieving promising clustering results. This limitation arises because k-means relies heavily on a single set of cluster centroids to represent the entire data space, which may be inadequate for capturing diverse patterns in real-world data. To overcome this limitation, several extensions of k-means have been proposed, including kernel-based variants, ensemble-based models, and multi-prototype formulations (Liu, Jiang, and Kot 2009; Liu et al. 2017; Bai, Liang, and Cao 2020; Guan and Terada 2023; Wang, Gittens, and Mahoney 2019; Nie, Wang, and Li 2019; Chen et al. 2021a; Jiang et al. 2024; Zeng et al. 2023; Li et al. 2023). For example, k-multiple means (Nie, Wang, and Li 2019) is a multi-prototype algorithm, which groups data samples with multiple sub-cluster means into specifed k clusters. Despite these efforts, many of these methods either rely on heuristic combinations of multiple clustering results or involve complex optimization procedures with significantly increased computational cost. The challenge remains in designing clustering methods that go beyond the simplistic assumptions of k-means to effectively capture complex data structures, while retaining computational efficiency and clustering accuracy.

Recently, multi-view clustering has demonstrated remarkable performance by using data with multiple views (Zheng et al. 2020; Fang et al. 2023; Zhang et al. 2024). Several k-means-based clustering methods under a multi-view set-



(a) K-means (ACC: 0.5060) (b) PMKC (ACC: 0.9690)

Figure 1: Visulization results of (a) k-means and (b) our proposed PMKC on a three circles dataset. We can observe that k-means fails to effectively group the data (ACC: 0.5060). In contrast, the proposed method achieves satisfactory clustering performance (ACC: 0.9690).

*Corresponding author

ting have also drawn increasing attention, with various algorithms proposed that achieve promising results (Zhang et al. 2025; Yang et al. 2022; Yang and Sinaga 2025). However, in many real-world scenarios, obtaining data with multiple views is usually difficult or even infeasible due to data collection limitations. Motivated by the success of multi-view clustering and aware of its practical constraints, we propose an interesting alternative: *instead of requiring data with multiple views to improve the k-means performance, we simulate a multi-view learning paradigm within a single-view setting*. In line with this motivation, we propose a framework termed Pseudo Multi-view K-means Clustering (PMKC), which can construct multiple soft k-means decompositions as distinct views of data samples to reveal different structural patterns.

To be specific, the underlying intuition of PMKC to simulates a multi-view learning paradigm is that a single clustering may fail to fully capture the complex intrinsic structure of real-world data, whereas multiple decompositions can provide complementary perspectives on the data distribution. This design enables PMKC to encode richer latent structures and reduce the risk of falling into suboptimal partitions caused by the local minima or biased centroids. Thus, each decomposition is regarded as a pseudo view, capturing a distinct structural perspective of the data. This is why we refer to our method as pseudo multi-view k-means clustering. To promote diversity among these pseudo views, we impose an independence constraint on the learned cluster centers, each decomposition to explore complementary structural aspects of the data. In parallel, to extract shared and consistent information across these decompositions, we also organize the soft assignment matrices into a third-order tensor and apply the low-rank tensor constraint. Overall, the designed framework enables PMKC to simultaneously capture complementary and consistent clustering information, resulting in more promising results. As illustrated in Fig. 1, PMKC can achieve superior performance on a three circles dataset, the structure of which is particularly challenging for conventional k-means. Meanwhile, PMKC also maintains a linear time cost with respective to the number of samples, which ensures its scalability to large datasets. Extensive experimental results verify the effectiveness and competitiveness of our PMKC, in comparison to several state-of-the-art clustering algorithms. Main contributions are as follows:

- We introduce a novel framework for enhancing the performance of k-means by simulating a multi-view learning paradigm. We generate multiple soft decompositions, each of which can be treated as a pseudo view of the data.
- We enforce independence among the cluster centers and impose a low-rank tensor constraint on soft assignments among pseudo views. The underlying structure information of sing-view data can be fully explored by PMKC.
- We design an optimization based on the augmented Lagrangian alternating direction minimization and achieve clustering results with a linear time cost with respect to the number of data samples. Experimental results demonstrate the superiority of the proposed method.

Related Work

In this section, we briefly review k-means and its extensions, followed by an introduction to multi-view clustering.

K-means and its Extensions

The classical k-means algorithm is one of the most widely used clustering methods due to its simplicity, efficiency, and ease of implementation (Ahmed, Seraj, and Islam 2020). It partitions data into k clusters by iteratively minimizing the sum of squared distances between data points and their corresponding cluster centroids. Nevertheless, k-means tends to yield less desirable clustering performance when applied to complex data, due to its inherent characteristics. To address this limitation, various approaches are introduced to extend k-means. One variant is kernel k-means (Likas, Vlassis, and Verbeek 2003; Dhillon, Guan, and Kulis 2004; Aradnia, Haeri, and Ebadzadeh 2022; Wang et al. 2024), which maps the input data into a high-dimensional feature space through a kernel function. This allows for capturing non-linear and complex cluster structures that are difficult to model using standard k-means. By operating in the kernel-induced space, kernel k-means can effectively separate clusters with arbitrary shapes, making it more suitable for datasets with nonlinearly separable patterns. However, its performance heavily depends on the choice of kernel (Nie et al. 2016; Yao et al. 2021). Another notable extension is Fuzzy c-means (FCM) (Bezdek, Ehrlich, and Full 1984; Hashemi, Gholian-Jouybari, and Hajiaghaei-Keshteli 2023), which relaxes the hard assignment constraint of classical k-means by allowing data points to belong to multiple clusters with varying degrees of membership. FCM introduces a membership matrix that assigns each sample a set of weights across all clusters, reflecting its degree of association with each cluster center. However, like k-means and its kernelized variant, FCM remains sensitive to initialization and may converge to suboptimal solutions. In addition to the above approaches, recent efforts also consider multi-prototype clustering (Nie, Wang, and Li 2019; Li et al. 2024; Guo et al. 2024), which improves upon classical k-means by representing each cluster with multiple prototypes instead of a single centroid.

Multi-view Clustering

Multi-view clustering has emerged as a powerful paradigm in unsupervised learning by leveraging complementary information from multiple feature representations or different modalities (Tang et al. 2020; Huang et al. 2023; Zhang et al. 2020; Fang et al. 2023; Zheng 2024; Yuan et al. 2025). Unlike traditional single-view methods such as k-means and its variants, multi-view clustering assumes that data are available from different "views", each capturing distinct characteristics of the underlying structure. By integrating information across views, these methods aim to achieve more robust and promising clustering performance.

Numerous multi-view clustering methods are proposed in recent years, including multi-view k-means clustering (Xu, Han, and Nie 2016; Yang et al. 2023; Yang and Sinaga 2025; Gao et al. 2025), which extends k-means to deal with multi-view data and demonstrates promising results in a variety of

applications. Nevertheless, most existing multi-view clustering methods require explicit access to multiple views, which may not be readily available in practice. In contrast, our proposed framework simulates a multi-view learning paradigm within a single-view setting, enabling richer modeling without relying on data with multiple views.

Methodology

Notations and Symbols

For clarity and without ambiguity, we adopt the following notational conventions: scalars are denoted by lowercase letters (e.g., a), vectors by bold lowercase letters (e.g., $\boldsymbol{a}$), matrices by bold uppercase letters (e.g., $\boldsymbol{A}$), and tensors by calligraphic letters (e.g., $\boldsymbol{\mathcal{A}}$). Regarding the 3-order tensor *e.g.*, $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor nuclear norm based on tensor-Singular Value Decomposition (t-SVD) (Kilmer et al. 2013; Lu et al. 2019) is denoted by $\|\boldsymbol{\mathcal{A}}\|_{\text{tensor}}$ as follows:

$$\|\boldsymbol{\mathcal{A}}\|_{\text{tensor}} = \frac{1}{n_3} \sum_{i=1}^{\min(n_1,n_2)} \sum_{j=1}^{n_3} g(\boldsymbol{\mathcal{S}}_f(i,i,j)), \quad (1)$$

where $\boldsymbol{\mathcal{S}}$ is obtained by conducting t-SVD on $\boldsymbol{\mathcal{A}}$, and $\boldsymbol{\mathcal{S}}_f$ denotes the Fourier transform of $\boldsymbol{\mathcal{S}}$. For more details, please refer to (Kilmer et al. 2013; Lu et al. 2019). By using different $g(\cdot)$, we can obtain different types of tensor nuclear norms (Xie et al. 2018; Guo et al. 2022; Chen et al. 2021b). Besides, we indicate the Hilbert Schmidt Independence Criterion (HSIC) between two matrices, e.g., $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$ and $\boldsymbol{B} \in \mathbb{R}^{n_1 \times n_2}$, as follows (Gretton et al. 2005):

$$\text{HSIC}(\boldsymbol{A}, \boldsymbol{B}) = (n_2 - 1)^{-2} \text{tr}(\boldsymbol{K_A L K_B L}), \quad (2)$$

where $\boldsymbol{K_A} \in \mathbb{R}^{n_2 \times n_2}$ and $\boldsymbol{K_B} \in \mathbb{R}^{n_2 \times n_2}$ are Gram matrices with the kernel function $\kappa(\cdot,\cdot)$, and $\boldsymbol{L} = \boldsymbol{I}_{n_2} - \frac{1}{n_2}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T$ is the centering matrix. We use the inner product kernel here.

Pseudo Multi-view K-means Clustering

We now present the proposed PMKC in detail. The core idea is to generate multiple soft k-means decompositions of the input data, treat each decomposition as an individual "view". Given a dataset $\boldsymbol{X} \in \mathbb{R}^{d \times n}$ with n samples and c clusters in d dimensional space, we have three key components:

1) Multiple soft k-means decompositions. We generate v soft k-means decompositions to simulate a multi-view learning process:

$$\begin{aligned} &\min_{\{\alpha_i, \boldsymbol{C}^{(i)}, \boldsymbol{Z}^{(i)}\}_{i=1}^v} \sum_{i=1}^{v} \frac{\alpha_i^2}{2} \|\boldsymbol{X} - \boldsymbol{C}^{(i)}\boldsymbol{Z}^{(i)}\|_F^2, \\ &\text{s.t.} \sum_{i=1}^{v} \alpha_i = 1,\ \alpha_i \geqslant 0,\ {\boldsymbol{Z}^{(i)}}^T \mathbf{1} = \mathbf{1},\ \boldsymbol{Z}^{(i)} \geqslant 0, \end{aligned} \quad (3)$$

where $\boldsymbol{C}^{(i)} \in \mathbb{R}^{d \times c}$ and $\boldsymbol{Z}^{(i)} \in \mathbb{R}^{c \times n}$ denote the matrices of cluster centers and soft assignments of the i-th soft k-means decompositions, α_i is the weight coefficient. Eq. (3) serves as the basis for constructing v pseudo views. Although each soft k-means decomposition can be viewed as exploring the underlying structure of the data from a different perspective, we should emphasize that PMKC do not have access to true multi-view data. If the decompositions are optimized based on Eq. (3) directly, they tend to converge to a similar solution, leading to highly redundant pseudo views. Therefore, to ensure the effectiveness of this pseudo multi-view design, it is necessary to introduce additional constraints that enforce structural diversity and facilitate meaningful integration across decompositions. To this end, we further impose independence constraints on the cluster centers and apply a low-rank tensor constraint over the tensor achieved by stacking assignment matrices.

2) Independence constraints on $\{\boldsymbol{C}^{(i)}\}_{i=1}^v$. To get that each pseudo view captures distinct aspects of the data, we impose independence constraints on matrices of cluster centers across different decompositions. In this work, we employ HSIC, which measures the statistical dependence between two random variables in a reproducing kernel Hilbert space (Gretton et al. 2005), as detailed below:

$$\min_{\{\boldsymbol{C}^{(i)}\}_{i=1}^v} \sum_{i=1}^{v} \sum_{j=1, j\neq i}^{v} \text{HSIC}(\boldsymbol{C}^{(i)}, \boldsymbol{C}^{(j)}), \quad (4)$$

which explicitly reduces the statistical dependence among pseudo views. The motivation behind Eq. (4) is that a smaller HSIC value indicates weaker dependency, implying greater independence between the two pseudo views. Intuitively, this encourages each clustering decomposition in Eq. (3) to construct cluster centers that capture orthogonal or uncorrelated structural patterns of the data and discourages degenerate solutions where different pseudo views converge to similar or overlapping partitions. Thus, the proposed PMKC is guided to explore a wider spectrum of underlying structures, increasing the representational richness of the overall clustering solution and reducing the risk of premature consensus across decompositions.

3) Low-rank tensor constraint. While encouraging the diversity among pseudo views is essential, it is equally important to extract a shared latent structure to produce a unified clustering result. To achieve this, we model the soft assignment matrices from each pseudo view as a third-order tensor and impose a low-rank regularization to promote cross-view consistency. Specifically, we have:

$$\min_{\boldsymbol{\mathcal{Z}}} \|\boldsymbol{\mathcal{Z}}\|_{\text{tensor}}, \text{ s.t. } \boldsymbol{\mathcal{Z}} = \varphi(\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}, \cdots, \boldsymbol{Z}^{(v)}), \quad (5)$$

where $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n \times v \times c}$ is the tensor achieved by stacking assignment matrices by $\varphi(\cdot)$. The motivation behind Eq. (5) stems from the assumption that although the pseudo views are diverse, they should share a common clustering structure embedded within their assignments. Since the low-rank tensor constraint has been extensively validated in multi-view clustering (Xie et al. 2018; Guo et al. 2022) as an effective strategy of capturing the consensus structure across diverse views, we adopt this constraint to guide the learning process of assignment matrices of different pseudo views.

Overall Objective Function. The overall objective of the proposed PMKC integrates three aforementioned key com-

ponents and can be formally written as follows:

$$\begin{aligned}&\min_{\{\alpha_i,\boldsymbol{C}^{(i)},\boldsymbol{Z}^{(i)}\}_{i=1}^{v}}\sum_{i=1}^{v}\frac{\alpha_i^2}{2}\|\boldsymbol{X}-\boldsymbol{C}^{(i)}\boldsymbol{Z}^{(i)}\|_F^2\\&+\lambda_1\sum_{i=1}^{v}\sum_{j=1,j\neq i}^{v}\mathrm{HSIC}(\boldsymbol{C}^{(i)},\boldsymbol{C}^{(j)})+\lambda_2\|\boldsymbol{\mathcal{Z}}\|_{\mathrm{tensor}},\\&\text{s.t.}\sum_{i=1}^{v}\alpha_i=1,\ \alpha_i\geqslant 0,\ {\boldsymbol{Z}^{(i)}}^T\mathbf{1}=\mathbf{1},\ \boldsymbol{Z}^{(i)}\geqslant 0,\\&\quad\boldsymbol{\mathcal{Z}}=\varphi(\boldsymbol{Z}^{(1)},\boldsymbol{Z}^{(2)},\cdots,\boldsymbol{Z}^{(v)}),\end{aligned}\tag{6}$$

where λ_1 and λ_2 are tradeoff hyperparameters.

Optimization

We present an augmented Lagrangian alternating direction algorithm (Lin, Liu, and Su 2011) for optimization. To make our minimization process separable, we introduce auxiliary variables $\boldsymbol{\mathcal{F}}=\boldsymbol{\mathcal{Z}}$ and $\{\boldsymbol{Q}^{(i)}=\boldsymbol{Z}^{(i)}\}_{i=1}^{v}$. Therefore, the following augmented Lagrange function $\boldsymbol{\mathcal{L}}$ can be achieved:

$$\begin{aligned}&\boldsymbol{\mathcal{L}}(\{\boldsymbol{C}^{(i)},\boldsymbol{Q}^{(i)},\boldsymbol{Z}^{(i)},\boldsymbol{Y}^{(i)},\alpha_i\}_{i=1}^{v},\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{W}})\\&=\sum_{i=1}^{v}\frac{\alpha_i^2}{2}\|\boldsymbol{X}-\boldsymbol{C}^{(i)}\boldsymbol{Q}^{(i)}\|_F^2\\&+\lambda_1\sum_{i=1}^{v}\sum_{j=1,j\neq i}^{v}\mathrm{HSIC}(\boldsymbol{C}^{(i)},\boldsymbol{C}^{(j)})+\lambda_2\|\boldsymbol{\mathcal{F}}\|_{\mathrm{tensor}},\\&+\sum_{i=1}^{v}\Psi(\boldsymbol{Y}^{(i)},\boldsymbol{Z}^{(i)}-\boldsymbol{Q}^{(i)})+\Psi(\boldsymbol{\mathcal{W}},\boldsymbol{\mathcal{Z}}-\boldsymbol{\mathcal{F}})\end{aligned}\tag{7}$$

where $\{\boldsymbol{Y}^{(i)}\}_{i=1}^{v}$ and $\boldsymbol{\mathcal{W}}$ indicate Lagrange multipliers. For $\varphi(\boldsymbol{A},\boldsymbol{B})$, it can be achieved as follows:

$$\varphi(\boldsymbol{A},\boldsymbol{B})=\frac{\mu}{2}\|\boldsymbol{B}\|_F^2+\langle\boldsymbol{A},\boldsymbol{B}\rangle,\tag{8}$$

where $\mu>0$ is a penalty variable. Subsequently, we iteratively optimize the following subproblems:

1) Optimizing $\boldsymbol{C}^{(i)}$-subproblem. We get the subproblem as follows:

$$\min_{\boldsymbol{C}^{(i)}}\sum_{i=1}^{v}\frac{\alpha_i^2}{2}\|\boldsymbol{X}-\boldsymbol{C}^{(i)}\boldsymbol{Q}^{(i)}\|_F^2+\sum_{j=1,j\neq i}^{v}\mathrm{HSIC}(\boldsymbol{C}^{(i)},\boldsymbol{C}^{(j)}).\tag{9}$$

By considering Eq. (2), we take the derivative with respect to $\boldsymbol{C}^{(i)}$ and set it to 0, the optimal solution is:

$$\boldsymbol{C}^{(i)}=\alpha_i^2\boldsymbol{X}{\boldsymbol{Q}^{(i)}}^T(\alpha_i^2\boldsymbol{Q}^{(i)}{\boldsymbol{Q}^{(i)}}^T+\frac{\lambda_1(\widetilde{\boldsymbol{L}}_i+\widetilde{\boldsymbol{L}}_i^T)}{(c-1)^2})^{-1},\tag{10}$$

where $\widetilde{\boldsymbol{L}}_i$ has the following definition with $\boldsymbol{L}=\boldsymbol{I}_c-\frac{1}{c}\mathbf{1}_c\mathbf{1}_c^T$:

$$\widetilde{\boldsymbol{L}}_i=\sum_{j=1,j\neq i}^{v}\boldsymbol{L}{\boldsymbol{C}^{(j)}}^T\boldsymbol{C}^{(j)}\boldsymbol{L}.\tag{11}$$

2) Optimizing $\boldsymbol{Q}^{(i)}$-subproblem. We update $\boldsymbol{Q}^{(i)}$ according to the following subproblem:

$$\min_{\boldsymbol{Q}^{(i)}}\sum_{i=1}^{v}\frac{\alpha_i^2}{2}\|\boldsymbol{X}-\boldsymbol{C}^{(i)}\boldsymbol{Q}^{(i)}\|_F^2+\Psi(\boldsymbol{Y}^{(i)},\boldsymbol{Z}^{(i)}-\boldsymbol{Q}^{(i)}).\tag{12}$$

Similar to the optimization of Eq. (9), we have:

$$\boldsymbol{Q}^{(i)}=(\alpha_i^2{\boldsymbol{C}^{(i)}}^T\boldsymbol{C}^{(i)}+\mu\boldsymbol{I})^{-1}(\mu\boldsymbol{Z}^{(i)}+\boldsymbol{Y}^{(i)}+\alpha_i^2{\boldsymbol{C}^{(i)}}^T\boldsymbol{X}).\tag{13}$$

3) Optimizing $\boldsymbol{Z}^{(i)}$-subproblem. To update $\boldsymbol{Z}^{(i)}$, we can solve the following subproblem:

$$\begin{aligned}&\min_{\boldsymbol{Z}^{(i)}}\Psi(\boldsymbol{Y}^{(i)},\boldsymbol{Z}^{(i)}-\boldsymbol{Q}^{(i)})+\Psi(\boldsymbol{W}^{(i)},\boldsymbol{Z}^{(i)}-\boldsymbol{F}^{(i)}),\\&\text{s.t. }{\boldsymbol{Z}^{(i)}}^T\mathbf{1}=\mathbf{1},\ \boldsymbol{Z}^{(i)}\geqslant 0,\end{aligned}\tag{14}$$

which can be further decomposed into n independent optimization subproblems. Specifically, for the k-th one, it has the following objective:

$$\min_{{\boldsymbol{z}_k^{(i)}}^T=1,\boldsymbol{z}_k^{(i)}\geqslant 0}\|\boldsymbol{z}_k^{(i)}-(\boldsymbol{q}_k^{(i)}+\boldsymbol{f}_k^{(i)}-\frac{\boldsymbol{y}_k^{(i)}+\boldsymbol{w}_k^{(i)}}{\mu})\|_2^2,\tag{15}$$

where $\boldsymbol{z}_k^{(i)}$, $\boldsymbol{q}_k^{(i)}$, $\boldsymbol{f}_k^{(i)}$, $\boldsymbol{y}_k^{(i)}$, and $\boldsymbol{w}_k^{(i)}$ are the k-th column of $\boldsymbol{Z}^{(i)}$, $\boldsymbol{Q}^{(i)}$, $\boldsymbol{F}^{(i)}$, $\boldsymbol{Y}^{(i)}$, and $\boldsymbol{W}^{(i)}$, respectively. The detailed optimization process of Eq. (15) is provided in the supplementary material.

4) Optimizing $\boldsymbol{\mathcal{F}}$-subproblem. By fixing other variables, we can achieve the following subproblem:

$$\min_{\boldsymbol{\mathcal{F}}}\frac{\lambda_2}{\mu}\|\boldsymbol{\mathcal{F}}\|_{\mathrm{tensor}}+\frac{1}{2}\|\boldsymbol{\mathcal{F}}-(\boldsymbol{\mathcal{Z}}+\frac{\boldsymbol{\mathcal{W}}}{\mu})\|_F^2.\tag{16}$$

The detailed optimization process of Eq. (15) is provided in the supplementary material. Regarding the type of tensor norm, we adopt the tensor nuclear norm (TNN) (Xie et al. 2018; Lu et al. 2019) by default in PMKC, which corresponds to setting $g(x)=x$ in Eq. (1). Furthermore, the influence of different tensor norm formulations on the clustering performance of PMKC is analyzed and discussed in detail in the experimental section.

5) Optimizing $\{\alpha_i\}_{i=1}^{v}$-subproblem. We have the following subproblem:

$$\min_{\sum_{i=1}^{v}\alpha_i=1,\alpha_i\geqslant 0}\sum_{i=1}^{v}\frac{\alpha_i^2}{2}\|\boldsymbol{X}-\boldsymbol{C}^{(i)}\boldsymbol{Q}^{(i)}\|_F^2.\tag{17}$$

By leveraging Cauchy-Schwarz inequality, we can update α_i with $t_i=\|\boldsymbol{X}-\boldsymbol{C}^{(i)}\boldsymbol{Q}^{(i)}\|_F$ as follows:

$$\alpha_i=\frac{1}{t_i}\Big/\sum\nolimits_{i=1}^{v}\frac{1}{t_i}.\tag{18}$$

6) Optimizing $\{\boldsymbol{Y}^{(i)},\boldsymbol{\mathcal{W}}\}$-subproblem. We can update Lagrange multipliers as follows:

Algorithm 1: Optimization of the proposed PMKC

Input: $\boldsymbol{X}$, λ_1, λ_2, c, v

Initialize $\mu = 10^{-6}$, $\rho = 2$, $\epsilon = 10^{-6}$, $\mu_{\max} = 10^6$, $\boldsymbol{Q}^{(i)} = \boldsymbol{0}$, $\boldsymbol{Z}^{(i)} = \boldsymbol{0}$, $\boldsymbol{Y}^{(i)} = \boldsymbol{0}$, $\boldsymbol{F}^{(i)} = \boldsymbol{0}$, $\boldsymbol{W}^{(i)} = \boldsymbol{0}$, and $\alpha_i = \frac{1}{v}$; Initialize $\boldsymbol{C}^{(i)}$ with random values.
while *not converged* **do**
  Update $\{\boldsymbol{Q}^{(i)}, \boldsymbol{C}^{(i)}, \boldsymbol{Z}^{(i)}, \alpha_i\}_{i=1}^{v}$, and $\boldsymbol{\mathcal{F}}$ according to *subproblems 1-5*;
  Update Lagrange multipliers and μ according to *subproblem 6*;
  Check the convergence conditions: $\|\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{Z}}\|_\infty < \epsilon$ and $\{\|\boldsymbol{Q}^{(i)} - \boldsymbol{Z}^{(i)}\|_\infty < \epsilon\}_{i=1}^{v}$,
end while
Output: $\{\boldsymbol{Z}^{(i)}\}_{i=1}^{v}$

$$\begin{cases} \boldsymbol{Y}^{(i)} = \boldsymbol{Y}^{(i)} + \mu(\boldsymbol{Z}^{(i)} - \boldsymbol{Q}^{(i)}), \\ \boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{W}} + \mu(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{F}}), \\ \mu = \min(\rho\mu, \mu_{\max}), \end{cases} \tag{19}$$

where $\rho > 1$ and $\mu_{\max}$ is the maximum value of μ.

Achieving clustering results. Upon convergence, we can obtain a set of soft assignment matrices $\{\boldsymbol{Z}^{(i)}\}_{i=1}^{v}$, each corresponding to a pseudo view. These matrices can be interpreted as anchor graphs constructed from cluster centroids, where each column of $\boldsymbol{Z}^{(i)}$ reflects the association strength between a data point and the anchor points (i.e., cluster centers) in the i-th view. To get final clustering results, we compute the average graph by taking the mean of all soft assignment matrices and employ a landmark-based spectral clustering algorithm (Chen and Cai 2011). For clarity, the overall optimization process is summarized in Algorithm 1.

Complexity and Convergence

We analyze the computational complexity of our PMKC. Let n be the number of samples, c the number of clusters, and v the number of pseudo views. In each iteration of optimization, the main computation burden can be composed of updating $\{\boldsymbol{C}^{(i)}, \boldsymbol{Q}^{(i)}, \boldsymbol{Z}^{(i)}, \boldsymbol{Y}^{(i)}, \alpha_i\}_{i=1}^{v}$, $\boldsymbol{\mathcal{F}}$, and $\boldsymbol{\mathcal{W}}$. For the i-th pseudo view, optimizing $\boldsymbol{C}^{(i)}$ and $\boldsymbol{Q}^{(i)}$ both require $\mathcal{O}(cn + c^3)$, updating $\boldsymbol{Z}^{(i)}$, $\boldsymbol{Y}^{(i)}$, and α_i all require $\mathcal{O}(cn)$, updating $\boldsymbol{\mathcal{F}}$ requires $\mathcal{O}(cnv(\log c + v))$, and updating $\boldsymbol{\mathcal{W}}$ requires $\mathcal{O}(cnv)$. Overall, the total complexity per iteration of the proposed optimization of our PMKC is:

$$\mathcal{O}(c^3 + cnv(\log c + v)), \tag{20}$$

which can be efficient in practice given that v and k are typically much smaller than n. Moreover, the use of landmark-based spectral clustering also requires $\mathcal{O}(n)$. Therefore, our PMKC inherits the desirable property of k-means in maintaining a linear time cost with respect to the number of samples, which enables it to efficiently deal with large-scale data scenarios commonly encountered in real-world applications.

Due to the presence of more than three subproblems, providing a strict convergence proof of Algorithm 1 is difficult (Xu 2021). Fortunately, our empirical results in the following section show that the optimization procedure converges rapidly and stably across diverse datasets.

Dataset	# Samples	# Dimensions	# Clusters
Breast-Cancer	669	9	2
WBC	683	9	2
Vowel	990	10	11
QSAR-Bio	1055	41	2
CMC	1473	9	3
Semeion	1593	256	10
Steel-Plates	1941	27	7
Segment	2310	19	7
Spambase	4601	57	2
Robot-Navigate	5456	24	4
Sat	6435	36	6
Magic	19020	10	2
Chess-KRVK	28056	6	18
Miniboone	130064	50	2

Table 1: Datasets information

Experiments

In this section, we conduct experiments to evaluate the effectiveness of the proposed PMKC method. Specifically, we begin by introducing the datasets and experimental settings, followed by a performance comparison with representative clustering algorithms. We then present detailed ablation results and provide discussions of our PMKC.

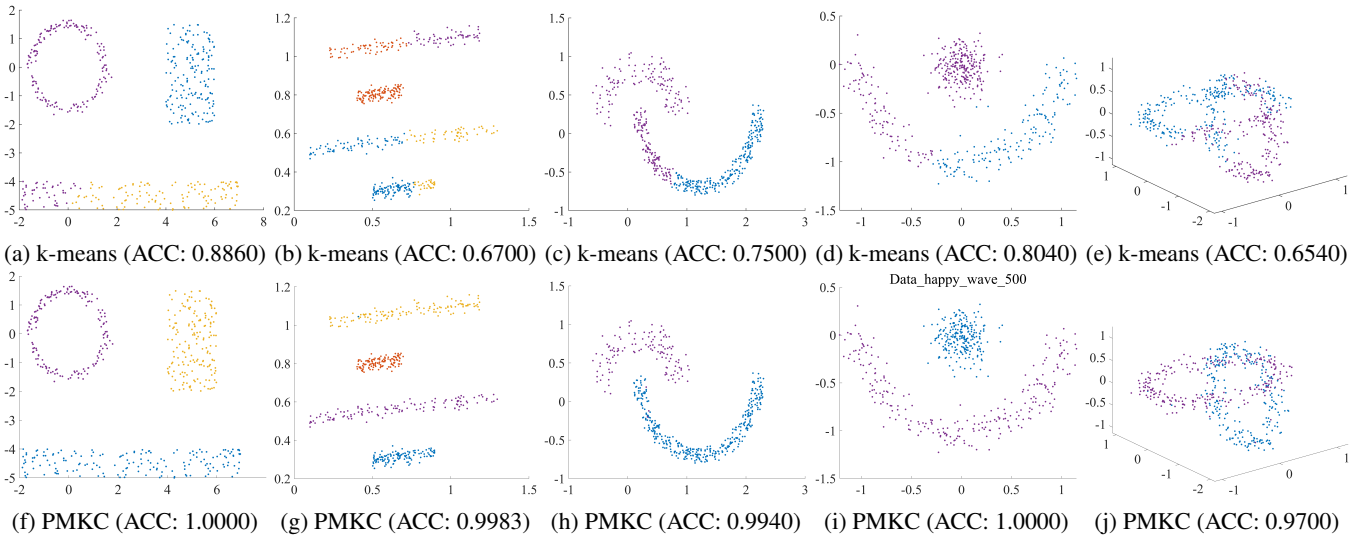
Experimental Setting

The datasets employed in our experiments are sourced from the UCI Machine Learning Repository[1]. A detailed summarization of these datasets is in Table 1. We compare PMKC with several representative clustering algorithms, including both classical and recent state-of-the-art methods: k-means, Spectral Clustering (SC) (Shi and Malik 2000), LRR (Liu et al. 2012), SSC (Elhamifar and Vidal 2013), LSR (Lu et al. 2012), LSpeC (Chen and Cai 2011), PCAN (Nie, Wang, and Huang 2014) , KMM (Nie, Wang, and Li 2019), FTRR (Ma et al. 2020), k-FSC (Fan 2021), FINCH (Sarfraz, Sharma, and Stiefelhagen 2019), LSubC (Kang et al. 2020), SGL (Kang et al. 2022), GCSC (Cai et al. 2020), AGCSC (Wei et al. 2023), and THBG (Zhao et al. 2025). We use the publicly available implementations and follow the hyperparameter settings suggested in the original works. For our PMKC, we select λ_1 and λ_2 from $\{0.001, 0.01, \ldots, 1000\}$, and fix $v = 3$. The clustering ACCuracy (ACC), Normalized Mutual Information (NMI), Purity, F-score, Precision, and running times in second are employed as evaluations to quantitatively depict the clustering performance. We run each experiment 30 times and report the mean of the results. Experimental environment is a Windows 11 computer with a 3.00 GHz Intel Xeon Gold 6154 CPU, and 64.0 GB RAM.

Clustering Performance

Visualization results on synthetic datasets. We conduct visualization experiments on synthetic datasets with com-

[1]https://archive.ics.uci.edu/



(a) k-means (ACC: 0.8860) (b) k-means (ACC: 0.6700) (c) k-means (ACC: 0.7500) (d) k-means (ACC: 0.8040) (e) k-means (ACC: 0.6540)

(f) PMKC (ACC: 1.0000) (g) PMKC (ACC: 0.9983) (h) PMKC (ACC: 0.9940) (i) PMKC (ACC: 1.0000) (j) PMKC (ACC: 0.9700)

Figure 2: Visulization results of (a)-(e) k-means and (f)-(j) our proposed PMKC on five synthetic datasets.

Dataset	k-means	SC	LRR	SSC	LSR	LSpeC	PCAN	KMM	FTRR	k-FSC	FINCH	LSubC	SGL	GCSC	AGCSC	THBG	PMKC
Breast-Cancer	0.937	0.954	0.936	0.936	0.949	0.944	0.760	0.955	0.944	0.627	0.877	0.939	0.939	0.947	0.950	0.950	**0.960**
WBC	0.943	0.959	0.939	0.940	0.952	0.968	0.703	0.956	0.965	0.588	0.952	0.940	0.940	0.949	0.956	0.895	**0.993**
Vowel	0.347	0.305	0.294	0.292	0.306	0.307	0.247	0.273	0.146	0.311	0.271	0.360	0.317	0.313	0.307	0.249	**0.383**
QSAR-Bio	0.528	0.509	0.619	0.551	0.623	0.541	0.509	0.596	0.652	0.529	0.629	0.528	0.530	0.672	0.617	0.531	**0.966**
CMC	0.383	0.386	0.411	0.395	0.417	0.407	0.355	0.401	0.418	0.396	0.393	0.388	0.391	0.421	0.409	0.367	**0.674**
Semeion	0.614	0.492	0.571	0.608	0.611	0.473	0.515	0.641	0.597	0.727	0.551	0.596	0.535	0.641	0.593	0.358	**0.925**
Steel-Pplates	0.409	0.409	0.419	0.412	0.440	0.386	0.376	0.416	0.492	0.419	0.447	0.406	0.398	0.433	0.452	0.441	**0.603**
Segment	0.677	0.706	0.722	0.714	0.721	0.604	0.570	0.520	0.613	0.518	0.640	0.657	0.682	0.731	0.729	0.408	**0.909**
Spambase	0.599	0.600	0.784	0.562	0.786	0.686	0.598	0.593	0.608	0.591	0.603	0.599	0.595	0.781	0.568	0.569	**0.901**
Robot-Navigate	0.405	0.377	0.373	0.368	0.368	0.364	0.437	0.432	0.411	0.406	0.446	0.376	0.374	0.375	0.384	0.435	**0.471**
Sat	0.695	0.669	0.653	0.536	0.746	0.610	0.272	0.706	0.706	0.677	0.662	0.717	0.730	0.748	0.712	0.533	**0.843**
Magic	0.590	0.608	0.674	0.559	0.557	0.639	0.513	0.649	0.599	0.725	0.699	0.585	0.562	0.551	0.627	0.719	**0.971**
Chess-KRVK	0.142	0.141	OM	OM	OM	0.137	0.189	0.202	OM	0.133	0.149	0.137	0.140	OM	OM	0.173	**0.356**
Miniboone	0.716	OM	OM	OM	OM	0.754	OM	OM	OM	0.726	OM	0.716	0.716	OM	OM	0.786	**0.812**

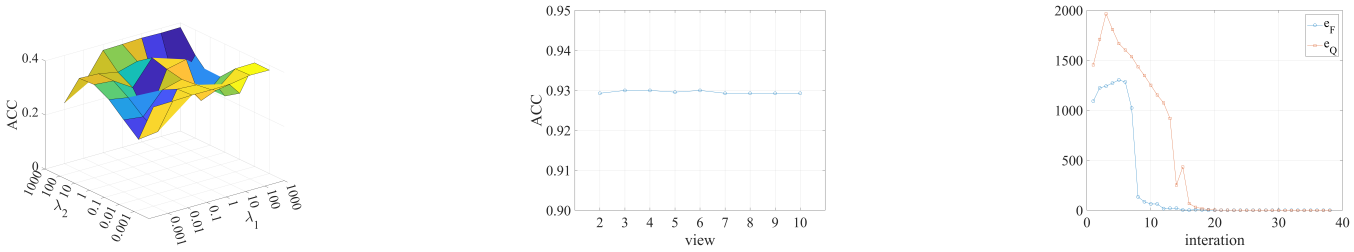
Table 2: Results in metric of ACC. The best results are highlighted in bold. "OM" denotes "Out-of-Memory error".

Dataset	k-means	SC	LRR	SSC	LSR	LSpeC	PCAN	KMM	FTRR	k-FSC	FINCH	LSubC	SGL	GCSC	AGCSC	THBG	PMKC
Breast-Cancer	0.654	0.721	0.651	0.649	0.694	0.727	0.160	0.718	0.704	0.032	0.493	0.659	0.662	0.683	0.729	0.703	**0.746**
WBC	0.683	0.747	0.659	0.665	0.708	0.783	0.082	0.730	0.769	0.063	0.716	0.672	0.672	0.694	0.752	0.543	**0.936**
Vowel	0.391	0.345	0.254	0.258	0.302	0.320	0.266	0.327	0.143	0.286	0.308	0.393	0.350	0.330	0.319	0.269	**0.484**
QSAR-Bio	0.078	0.088	0.079	0.119	0.077	0.051	0.085	0.041	0.007	0.099	0.020	0.078	0.077	0.102	0.066	0.079	**0.773**
CMC	0.032	0.029	0.022	0.021	0.023	0.026	0.018	0.009	0.008	0.020	0.017	0.029	0.028	0.024	0.027	0.012	**0.294**
Semeion	0.556	0.454	0.451	0.542	0.495	0.402	0.581	0.659	0.602	0.688	0.601	0.529	0.515	0.598	0.453	0.375	**0.914**
Steel-Plates	0.285	0.301	0.313	0.302	0.328	0.260	0.272	0.294	0.339	0.303	0.319	0.284	0.285	0.311	0.330	0.278	**0.658**
Segment	0.622	0.659	0.625	0.651	0.625	0.559	0.687	0.652	0.612	0.461	0.689	0.648	0.651	0.623	0.666	0.413	**0.916**
Spambase	0.010	0.008	0.233	0.054	0.254	0.144	0.012	0.020	0.004	0.002	0.005	0.010	0.015	0.252	0.047	0.046	**0.620**
Robot-Navigate	0.118	0.111	0.112	0.077	0.077	0.087	0.043	0.097	0.030	0.082	0.061	0.112	0.107	0.081	0.079	0.133	**0.134**
Sat	0.633	0.562	0.515	0.418	0.594	0.498	0.035	0.654	0.579	0.562	0.536	0.577	0.588	0.595	0.579	0.388	**0.854**
Magic	0.017	0.029	0.093	0.008	0.007	0.066	0.000	0.001	0.000	0.117	0.115	0.016	0.009	0.005	0.045	0.103	**0.816**
Chess-KRVK	0.127	0.118	OM	OM	OM	0.122	0.129	0.125	OM	0.092	0.105	0.119	0.126	OM	OM	0.074	**0.557**
Miniboone	0.003	OM	OM	OM	OM	0.149	OM	OM	OM	0.026	OM	0.003	0.003	OM	OM	0.172	**0.211**

Table 3: Results in metric of NMI. The best results are highlighted in bold. "OM" denotes "Out-of-Memory error".

plex cluster structures to intuitively demonstrate the clustering behavior of our proposed PMKC method in comparison with the k-means algorithm. As depicted in Fig. 2, PMKC can produce significantly more coherent and well-separated clusters than k-means, especially in cases where the data exhibits complex structures. This qualitative comparison highlights the superior capability of PMKC in capturing the underlying structure of the data.

Quantitative results on real-world datasets. We present a quantitative evaluation of PMKC and comparison methods on datasets listed in Table 1. As shown in Tables 2 and 3, PMKC achieves the best performance in terms of ACC and NMI across all datasets. This superior performance can be attributed to the mechanism of simulating multi-view learning in PMKC, which generates pseudo multiple views and excavates information from different perspectives, thereby



(a) (b) (c)

Figure 3: Experimental results of PMKC on Breast-Cancer dataset. (a) and (b) provide results with different hyperparameters, (c) shows convergence curves.

better revealing the underlying structure of the data. As a result, PMKC delivers more promising clustering results compared to existing state-of-the-art methods. Moreover, we report the running-time comparison in appendix, where PMKC demonstrates competitive efficiency.

Model Analysis

1) Parameter sensitivity and convergence. Three hyper-parameters, namely, λ_1, λ_2, and v, are involved in PMKC. We take the Breast-Cancer dataset as an example and present the results in Fig. 3(a)–(b). The promising clustering ACC results can be obtained by PMKC within larger relatively ranges of λ_1 and λ_2. Meanwhile, it can be observed that PMKC maintains consistently good ACC results with $v \geqslant 2$, demonstrating its effectiveness in simulating multi-view learning paradigm. As illustrated in Fig. 3 (c), we provide curves of $e_F = \|\boldsymbol{\mathcal{F}} - \boldsymbol{\mathcal{Z}}\|_\infty$ and $e_Q = \sum_i \|\boldsymbol{Q}^{(i)} - \boldsymbol{Z}^{(i)}\|_\infty$. It can be observed that our optimization algorithm has a stable and fast convergence property. A theoretical convergence analysis will be explored in our future works.

2) Impact of different tensor norms. To show the influence of adopting different tensor nuclear norms, we further consider using Geman tensor norm (Chen et al. 2021b), i.e., $g(x) = \frac{(1+\theta)x}{\theta+x}$ with $\theta = 2$, and logarithmic tensor norm (Guo et al. 2022), i.e., $g(a) = \log(1 + a^p)$ with $p = 1$, in Eq. (1). To be clear, we denote PMKC with Geman tensor norm by PMKC$_G$ and logarithmic tensor norm by PMKC$_L$. As shown Table 4, promising results can be achieved in all cases. This observation indicates that the proposed PMKC is robust to the types of tensor nuclear norm, which is attractive to employing the our method in reallife applications.

3) Ablation studies. We indicate the method that ignores the independent constraints in PMKC by PMKC-w/o-IC and the method that ignores the low-rank tensor constraint by PMKC-w/o-LTC. The clustering results are provided in Table 5. It can be observed that PMKC outperforms PMKC-w/o-IC and PMKC-w/o-LTC in all cases, which verified the effectiveness of employing both independent constraints and low-rank tensor constraints in the proposed method.

Datasets	PMKC		PMKC$_G$		PMKC$_L$	
	ACC	NMI	ACC	NMI	ACC	NMI
Breast-Cancer	0.9599	0.7460	0.9614	0.7501	0.9557	0.7273
WBC	0.9927	0.9360	0.9926	0.9413	0.9912	0.9237
Vowel	0.3825	0.4836	0.3447	0.3654	0.5154	0.6544
QSAR-Bio	0.9659	0.7733	0.9498	0.7372	0.9706	0.7958
CMC	0.6743	0.2939	0.7488	0.3839	0.7034	0.5087
Semeion	0.9249	0.9144	0.9120	0.8713	0.8999	0.8848
Steel-Plates	0.6026	0.6575	0.5484	0.5046	0.6252	0.6023
Segment	0.9093	0.9158	0.9232	0.8920	0.9103	0.9154
Spambase	0.9005	0.6195	0.9037	0.6273	0.9015	0.6221
Robot-Navigate	0.4711	0.1340	0.4493	0.1448	0.4558	0.1460
Sat	0.8429	0.8543	0.8166	0.7258	0.8048	0.7571
Magic	0.9705	0.8162	0.8097	0.3055	0.9903	0.9269
Chess-KRVK	0.3557	0.5567	0.3326	0.5118	0.2543	0.3767
Miniboone	0.8123	0.2112	0.9188	0.6400	0.9467	0.7260

Table 4: Results of PMKC, PMKC$_G$, and PMKC$_L$.

Datasets	PMKC-w/o-IC		PMKC-w/o-LTC		PMKC	
	ACC	NMI	ACC	NMI	ACC	NMI
Breast-Cancer	0.9299	0.6352	0.9156	0.5884	0.9599	0.7460
WBC	0.9488	0.7050	0.9272	0.6274	0.9927	0.9360
Vowel	0.3328	0.3955	0.2982	0.2839	0.3825	0.4836
QSAR-Bio	0.6616	0.0012	0.5166	0.0862	0.9659	0.7733
CMC	0.3992	0.0396	0.3993	0.0355	0.6743	0.2939
Semeion	0.7170	0.6628	0.5029	0.4495	0.9249	0.9144
Steel-Plates	0.5123	0.4325	0.4158	0.2743	0.6026	0.6575
Segment	0.7798	0.7749	0.6542	0.5894	0.9093	0.9158
Spambase	0.7885	0.2460	0.7881	0.2535	0.9005	0.6195
Robot-Navigate	0.3901	0.0936	0.3948	0.0938	0.4711	0.1340
Sat	0.6699	0.6674	0.7164	0.5713	0.8429	0.8543
Magic	0.6646	0.0253	0.6310	0.0059	0.9705	0.8162
Chess-KRVK	0.2476	0.3504	0.1507	0.1220	0.3557	0.5567
Miniboone	0.7161	0.0574	0.7507	0.0976	0.8123	0.2112

Table 5: Results of ablation studies.

Conclusion

In this paper, we proposed a Pseudo Multi-view K-means Clustering (PMKC) method that processes single-view data by simulating a multi-view learning paradigm. By imposing independent constraints on cluster centers and and employing low-rank tensor constraint on soft assignments of different pseudo views, PMKC can better capture the underlying data structure, leading to more promising clustering performance. Experimental results on multiple real-world and synthetic datasets demonstrate that PMKC consistently outperforms state-of-the-art methods.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62306074, 62206156), the Natural Science Foundation of Fujian Province (2023J05025).

References

Ahmed, M.; Seraj, R.; and Islam, S. M. S. 2020. The k-means algorithm: a comprehensive survey and performance evaluation. *Electronics*, 9(8): 1295.

Aradnia, A.; Haeri, M. A.; and Ebadzadeh, M. M. 2022. Adaptive explicit kernel Minkowski weighted k-means. *Appl. Math. Comput.*, 584: 503–518.

Arthur, D.; and Vassilvitskii, S. 2006. k-means++: the advantages of careful seeding. Technical report, Stanford.

Bai, L.; Liang, J.; and Cao, F. 2020. A multiple k-means clustering ensemble algorithm to find nonlinearly separable clusters. *Inf. Fusion*, 61: 36–47.

Bezdek, J. C.; Ehrlich, R.; and Full, W. 1984. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.*, 10(2): 191–203.

Cai, Y.; Zhang, Z.; Cai, Z.; Liu, X.; Jiang, X.; and Yan, Q. 2020. Graph convolutional subspace clustering: a robust subspace clustering framework for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.*, 59(5): 4191–4202.

Chen, L.; Zhou, S.; Ma, J.; and Xu, M. 2021a. Fast kernel k-means clustering using incomplete Cholesky factorization. *Appl. Math. Comput.*, 402: 126037.

Chen, X.; and Cai, D. 2011. Large scale spectral clustering with landmark-based representation. In *AAAI*, volume 25, 313–318.

Chen, Y.; Wang, S.; Peng, C.; Hua, Z.; and Zhou, Y. 2021b. Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering. *IEEE Trans. Image Process.*, 30: 4022–4035.

Dhillon, I. S.; Guan, Y.; and Kulis, B. 2004. Kernel k-means: spectral clustering and normalized cuts. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 551–556.

Elhamifar, E.; and Vidal, R. 2013. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(11): 2765–2781.

Fan, J. 2021. Large-scale subspace clustering via k-factorization. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 342–352.

Fang, U.; Li, M.; Li, J.; Gao, L.; Jia, T.; and Zhang, Y. 2023. A comprehensive survey on multi-view clustering. *IEEE Trans. Knowl. Data Eng.*, 35(12): 12350–12368.

Gao, Q.; Li, F.; Wang, Q.; Gao, X.; and Tao, D. 2025. Manifold based multi-view k-means. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(4): 3175–3182.

Gretton, A.; Bousquet, O.; Smola, A.; and Schölkopf, B. 2005. Measuring statistical dependence with Hilbert-Schmidt norms. In *Int. Conf. Algorithmic Learn. Theory*, 63–77. Springer.

Guan, X.; and Terada, Y. 2023. Sparse kernel k-means for high-dimensional data. *Pattern Recognition*, 144: 109873.

Guo, J.; Sun, Y.; Gao, J.; Hu, Y.; and Yin, B. 2022. Logarithmic schatten-p p norm minimization for tensorial multi-view subspace clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(3): 3396–3410.

Guo, M.-H.; Zhang, Y.; Mu, T.-J.; Huang, S. X.; and Hu, S.-M. 2024. Tuning vision-language models with multiple prototypes clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12): 11186–11199.

Hashemi, S. E.; Gholian-Jouybari, F.; and Hajiaghaei-Keshteli, M. 2023. A fuzzy c-means algorithm for optimizing data clustering. *Expert Syst. Appl.*, 227: 120377.

Huang, S.; Tsang, I. W.; Xu, Z.; and Lv, J. 2023. Latent representation guided multi-view clustering. *IEEE Trans. Knowl. Data Eng.*, 35(7): 7082–7087.

Jiang, S. H.-C.; Krauthgamer, R.; Lou, J.; and Zhang, Y. 2024. Coresets for kernel clustering. *Machine Learning*, 113(8): 5891–5906.

Kang, Z.; Lin, Z.; Zhu, X.; and Xu, W. 2022. Structured graph learning for scalable subspace clustering: from single view to multiview. *IEEE Trans. Cybern.*, 52(9): 8976–8986.

Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale multi-view subspace clustering in linear time. In *AAAI*, volume 34, 4412–4419.

Kilmer, M. E.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.*, 34(1): 148–172.

Li, D.; Kosugi, S.; Zhang, Y.; Okumura, M.; Xia, F.; and Jiang, R. 2025. Revisiting dynamic graph clustering via matrix factorization. In *Proc. ACM on Web Conf.*, 1342–1352.

Li, D.; Zhou, S.; Zeng, T.; and Chan, R. H. 2023. Multi-prototypes convex merging based k-means clustering algorithm. *IEEE Trans. Knowl. Data Eng.*, 36(11): 6653–6666.

Li, D.; Zhou, S.; Zeng, T.; and Chan, R. H. 2024. Multi-prototypes convex merging based k-means clustering algorithm. *IEEE Trans. Knowl. Data Eng.*, 36(11): 6653–6666.

Likas, A.; Vlassis, N.; and Verbeek, J. J. 2003. The global k-means clustering algorithm. *Pattern Recognition*, 36(2): 451–461.

Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Adv. Neural Inform. Process. Syst.*, volume 24, 1–9. Curran Associates, Inc.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2012. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1): 171–184.

Liu, H.; Wu, J.; Liu, T.; Tao, D.; and Fu, Y. 2017. Spectral ensemble clustering via weighted k-means: Theoretical and practical evidence. *IEEE Trans. Knowl. Data Eng.*, 29(5): 1129–1143.

Liu, M.; Jiang, X.; and Kot, A. C. 2009. A multi-prototype clustering algorithm. *Pattern Recognition*, 42(5): 689–698.

Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(4): 925–938.

Lu, C.-Y.; Min, H.; Zhao, Z.-Q.; Zhu, L.; Huang, D.-S.; and Yan, S. 2012. Robust and efficient subspace segmentation

via least squares regression. In *Eur. Conf. Comput. Vis.*, 347–360. Springer.

Ma, Z.; Kang, Z.; Luo, G.; Tian, L.; and Chen, W. 2020. Towards clustering-friendly representations: Subspace clustering via graph filtering. In *ACM Int. Conf. Multimedia*, 3081–3099.

Nie, F.; Wang, C.-L.; and Li, X. 2019. K-multiple-means: a multiple-means clustering method with specified k clusters. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 959–967.

Nie, F.; Wang, X.; and Huang, H. 2014. Clustering and projected clustering with adaptive neighbors. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 977–986.

Nie, F.; Wang, X.; Jordan, M.; and Huang, H. 2016. The constrained laplacian rank algorithm for graph-based clustering. In *AAAI*, volume 30.

Oyewole, G. J.; and Thopil, G. A. 2023. Data clustering: application and trends. *Machine Learning*, 56(7): 6439–6475.

Ren, Z.; Sun, Q.; and Wei, D. 2021. Multiple kernel clustering with kernel k-means coupled graph tensor learning. In *AAAI*, volume 35, 9411–9418.

Sarfraz, S.; Sharma, V.; and Stiefelhagen, R. 2019. Efficient parameter-free clustering using first neighbor relations. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Shi, J.; and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8): 888–905.

Tang, C.; Liu, X.; Zhu, X.; Zhu, E.; Luo, Z.; Wang, L.; and Gao, W. 2020. CGD: multi-view clustering via cross-view graph diffusion. In *AAAI*, volume 34, 5924–5931.

Von Luxburg, U. 2007. A tutorial on spectral clustering. *Stat. Comput.*, 17(4): 395–416.

Wang, J.; Tang, C.; Zheng, X.; Liu, X.; Zhang, W.; Zhu, E.; and Zhu, X. 2024. Fast approximated multiple kernel k-Means. *IEEE Trans. Knowl. Data Eng.*, 36(11): 6171–6180.

Wang, S.; Gittens, A.; and Mahoney, M. W. 2019. Scalable kernel k-means clustering with nystrom approximation: Relative-error bounds. *J. Mach. Learn. Res.*, 20(12): 1–49.

Wei, L.; Chen, Z.; Yin, J.; Zhu, C.; Zhou, R.; and Liu, J. 2023. Adaptive graph convolutional subspace clustering. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 6262–6271.

Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *Int. J. Comput. Vis.*, 126: 1157–1179.

Xu, J.; Han, J.; and Nie, F. 2016. Discriminatively embedded k-means for multi-view clustering. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Xu, Y. 2021. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. *Math. Program.*, 185(1): 199–244.

Yang, B.; Zhang, X.; Li, Z.; Nie, F.; and Wang, F. 2022. Efficient multi-view K-means clustering with multiple anchor graphs. *IEEE Trans. Knowl. Data Eng.*, 35(7): 6887–6900.

Yang, B.; Zhang, X.; Li, Z.; Nie, F.; and Wang, F. 2023. Efficient multi-view k-Means clustering with multiple anchor graphs. *IEEE Trans. Knowl. Data Eng.*, 35(7): 6887–6900.

Yang, M.-S.; and Sinaga, K. P. 2025. Federated multi-view K-means clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(4): 2446–2459.

Yao, Y.; Li, Y.; Jiang, B.; and Chen, H. 2021. Multiple kernel k-Means clustering by selecting representative kernels. *IEEE Trans. Neural Netw. Learn. Syst.*, 32(11): 4983–4996.

Yin, J.; Sun, S.; Wei, L.; and Wang, P. 2024. Discriminatively fuzzy multi-view k-means clustering with local structure preserving. In *AAAI*, volume 38, 16478–16485.

Yuan, H.; Sun, Y.; Zhou, F.; Wen, J.; Yuan, S.; You, X.; and Ren, Z. 2025. Prototype matching learning for incomplete multi-view clustering. *IEEE Trans. Image Process.*, 34: 828–841.

Zeng, S.; Duan, X.; Bai, J.; Tao, W.; Hu, K.; and Tang, Y. 2023. Soft multiprototype clustering algorithm via two-layer semi-nmf. *IEEE Trans. Fuzzy Syst.*, 32(4): 1615–1629.

Zhang, C.; Fu, H.; Hu, Q.; Cao, X.; Xie, Y.; Tao, D.; and Xu, D. 2020. Generalized latent multi-view subspace clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(1): 86–99.

Zhang, C.; Jia, X.; Li, Z.; Chen, C.; and Li, H. 2024. Learning cluster-wise anchors for multi-view clustering. In *AAAI*, volume 38, 16696–16704.

Zhang, Z.; Chen, X.; Wang, C.; Wang, R.; Song, W.; and Nie, F. 2025. Structured multi-view k-means clustering. *Pattern Recognition*, 160: 111113.

Zhao, Z.; Cao, Z.; Xin, H.; Wang, R.; Wu, D.; Wang, Z.; and Nie, F. 2025. Enhancing clustering performance with tensorized high-order bipartite graphs: a structured graph learning approach. *IEEE Trans. Circuit Syst. Video Technol.*, 35(3): 2616–2631.

Zheng, Q. 2024. Flexible and parameter-free graph learning for multi-view spectral clustering. *IEEE Trans. Circuit Syst. Video Technol.*, 34(9): 8966–8971.

Zheng, Q.; Zhu, J.; Li, Z.; Pang, S.; Wang, J.; and Li, Y. 2020. Feature concatenation multi-view subspace clustering. *Neurocomputing*, 379: 89–102.