Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

From Individual to Universal: Regularized Multi-view Joint Representation for Multi-view Subspace-Preserving Recovery

Libin Wang[1], **Yulong Wang**[1,2,*], **Xinwei He**[1,2,*], **Qiwei Xie**[3], **Kit Ian Kou**[4] and **Yuan Yan Tang**[5]
[1]College of Informatics, Huazhong Agricultural University, China
[2]Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education, China
[3]Beijing Smarter Eye Technology Co., Ltd
[4]Faculty of Science and Technology, University of Macau, Macau, China
[5]Zhuhai UM Science and Technology Research Institute, FST, University of Macau, Macau, China
wlb123@webmail.hzau.edu.cn, wangyulong6251@gmail.com, xwhe@mail.hzau.edu.cn,
qiwei.xie@bjut.edu.cn, {kikou, yytang}@umac.mo

Abstract

Recent years have witnessed an explosion of Multi-view Subspace Classification (MSCla) and Multi-view Subspace Clustering (MSClu) methods for various applications. However, their theoretical foundation have not been well explored and understood. In this paper, we investigate the multi-view subspace-preserving recovery theory, which is the theoretical underpinnings for MSCla and MSClu methods. Specifically, we derive novel geometrically interpretable conditions for the success of multi-view subspace-preserving recovery. Compared with prior related works, we make the following innovations: *First*, our theory does not require the equality constraint, which is a common requirement in prior theoretical works and may be too restrictive in reality. *Second*, we provide both Individual Theoretical Guarantee (ITG) and Universal Theoretical Guarantee (UTG) for multi-view subspace-preserving recovery while prior works only give the UTG. *Third*, we also apply the proposed theory to establish theoretical guarantees for MSCla and MSClu, respectively. Numerical results validate the proposed theory for multi-view subspace-preserving recovery.

1 Introduction

High-dimensional data in multiple classes can often be modeled as samples drawn from a union of low-dimensional subspaces in many problems of machine learning [Wright *et al.*, 2009; Vidal, 2011]. For example, motion trajectories in a video [Costeira and Kanade, 1998], movie ratings [Zhang *et al.*, 2012], and facial images under varying illumination [Basri and Jacobs, 2003] can be approximately represented by subspaces, with each subspace corresponding to a class. Numerous methods of subspace classification [Wright *et al.*, 2009; Zhang *et al.*, 2022] and subspace clustering [Elhamifar and Vidal, 2013; Li *et al.*, 2017] have been developed to capture the low-dimensional structure of high-dimensional data in recent years. Subspace based methods have also been employed in many applications such as face recognition [Zhang *et al.*, 2024], motion segmentation [Ma *et al.*, 2008], and cancer subtype clustering [Li *et al.*, 2017], etc.

In spite of the effectiveness of these methods, they only involve the single-view data. In many real-world applications, data is often multi-view in nature. For example, data of web pages usually contain hyperlinks, texts and visual information. To better exploit the consensus and complementary information of multi-view data, several Multi-view Subspace Classification (MSCla) [Shekhar *et al.*, 2013; Yang *et al.*, 2024] and Multi-view Subspace Clustering (MSClu) [Zhang *et al.*, 2020; Chang *et al.*, 2024] methods have been proposed recently. For instance, [Wang *et al.*, 2023] proposes a novel multi-view sparse subspace clustering method based on the joint sparse representation (JSR).

Despite the empirical success of the works above, their theoretical foundations have not been well explored. For instance, it is not fully understood under what conditions these methods succeed in classification or clustering. To understand the correctness of them, several theoretical advances have been developed, which are committed to establishing the *Subspace-Preserving Recovery* [You and Vidal, 2015].

1.1 Single-view Subspace-Preserving Recovery

Due to its fundamental role, subspace-preserving recovery has attracted much attention in the theoretical analysis of subspace classification and clustering [You *et al.*, 2016; Wang *et al.*, 2019]. Given a matrix $\mathbf{D} \in \mathbb{R}^{d\times N}$ with its columns from the union of multiple subspaces and a data point $\mathbf{y} \in \mathbb{R}^d$ from one of these subspaces, subspace-preserving recovery aims to seek a subspace-preserving representation (SPR) $\mathbf{c} \in \mathbb{R}^N$ of $\mathbf{y}$ by solving $\mathbf{y} = \mathbf{D}\mathbf{c}$. This SPR ensures that the nonzero entries of $\mathbf{c}$ conform with the columns of $\mathbf{D}$ that belong to the same subspace as $\mathbf{y}$ [You and Vidal, 2015].

Many previous works study the theory of how to yield a SPR in the subspace clustering and classification problems. [Elhamifar and Vidal, 2009; Elhamifar and Vidal, 2010] establish the SPR guarantee for the Basis Pursuit (BP) model by

*Corresponding authors.

assuming that the subspaces are independent and disjoint, respectively. [Soltanolkotabi and Candés, 2012] broadens prior results for the sparse subspace clustering when the subspaces have nontrivial intersections. [You and Vidal, 2015] provides SPR guarantees for both BP and OMP (Orthogonal Matching Pursuit). Besides, [Wang *et al.*, 2019] extends the SPR theory to the atomic representation model for subspace classification. More prior works can be explored in [Wang and Xu, 2016; Kaba *et al.*, 2021; Thaker *et al.*, 2022].

1.2 Multi-view Subspace-Preserving Recovery

Despite enhancing our comprehension, all the results above are confined to single-view data and this limits their applications in real-world problems. To extend the SPR theory to multi-view data, [Wang, 2024] considers the multi-view joint sparse representation (MJSR) model

$$\min_{\mathbf{C}\in\mathbb{R}^{N\times V}} \|\mathbf{C}\|_{1,2}, \text{ s.t. } \mathbf{y}^v = \mathbf{D}^v\mathbf{C}(:,v),\ v=1,\cdots,V, \quad (1)$$

where $\mathbf{y}^v$ and $\mathbf{D}^v$ denote the test sample and the data matrix of the v-th view, respectively. Here, $\|\mathbf{C}\|_{1,2} = \sum_{i=1}^m \|\mathbf{C}(i,:)\|_2$ is the matrix $\ell_{1,2}$ norm of the representation matrix $\mathbf{C}$ to encourage row-sparse solutions and thus exploits the correlation information among different views. [Wang, 2024] proposes the atomic recovery property (ARP) and proves that the solution of MJSR model is a Multi-view Subspace-Preserving Representation (MSPR) when the ARP holds.

However, the work [Wang, 2024] is confined to the equality constrained MJSR model, which may be too strict for practical use. The analysis of the multi-view subspace-preserving recovery theory of *regularized MJSR model* is still missing in the existing literature, which is more flexible in reality. Moreover, most existing theoretical results only provide the Universal Theoretical Guarantee (UTG), which require strong assumptions for all samples in corresponding subspaces and may break down when part of samples fail to satisfy these assumptions. Thus, the theoretical analysis *individually established for a fixed data point* is also missing. To fill these gaps, in this paper we propose new theoretical conditions and results with *concise geometric characterizations*.

1.3 Paper Contributions

In this work, we consider the Regularized Multi-view Joint Sparse Representation (RMJSR) model

$$\min_{\mathbf{C}\in\mathbb{R}^{n\times V}} \|\mathbf{C}\|_{1,2} + \frac{\gamma}{2}\sum_{v=1}^{V} \|\mathbf{y}^v - \mathbf{D}^v\mathbf{C}(:,v)\|_2^2, \quad (2)$$

where γ denotes the regularization parameter. Let $\mathbf{C}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}})$ denote any optimal solution to problem (2) (The descriptions of $\dot{\mathbf{y}}$ and $\dot{\mathbf{D}}$ can be seen in Tab. 1). Compared with MJSR, RMJSR does not require the equality constraints in Eq. (1), which do not necessarily hold in reality. In this work, we investigate theoretical conditions under which the solution of the RMJSR model is a MSPR (See Definition 1). Specifically, the contributions of this work can be summarized as follows.

1. *Theoretical guarantee under milder assumptions.* We establish the theoretical guarantees of RMJSR for multi-view subspace-preserving recovery. Our theory does not require the equality constraint, which is a common requirement in prior theoretical results and may be too restrictive in reality. To the best of the authors' knowledge, this is the first theoretical work of the multi-view subspace-preserving recovery theory for RMJSR.
2. *Theory for both individual and universal cases.* We provide both Individual Theoretical Guarantee (ITG) and Universal Theoretical Guarantee (UTG) for MSPR while the previous works only give the UTG. Specifically, we derive theoretical conditions under which RMJSR succeeds in yielding MSPR for a specific individual test sample and all samples, respectively.
3. *Applications on MSCla and MSClu.* Based on the proposed theory, we derive novel theoretical guarantees for multi-view subspace classification and multi-view subspace clustering, respectively.

2 Preliminaries

2.1 Notations

To improve the readability, we denote scalars, vectors, matrices and sets as italic letters (e.g., a), boldface lowercase letters (e.g., $\mathbf{a}$), boldface capital letters (e.g., $\mathbf{A}$) and calligraphic capital letters (e.g., $\mathcal{A}$), respectively. For any vector $\mathbf{a} \in \mathbb{R}^N$, supp($\mathbf{a}$) denotes its support set, i.e., $\text{supp}(\mathbf{a}) = \{i : a_i \neq 0\}$. Similarly, for any matrix $\mathbf{A}$, rowsupp($\mathbf{A}$) denotes its row support set, i.e., the index set of nonzero rows of $\mathbf{A}$.

To distinguish multi-view data from single-view data, a dot is added on the above of the variable (e.g., $\dot{\mathbf{a}}$). Tab. 1 summarizes the key notations and acronyms used in this paper.

2.2 Multi-view Subspace-Preserving Recovery

To extend the SPR theory to the multi-view data, [Wang, 2024] tackles the multi-view joint sparse representation (MJSR) model in Eq. (1) and gives the definition of the multi-view subspace-preserving representation (MSPR) as follows.

Definition 1. ***(Multi-view Subspace-Preserving Representation, MSPR)*** *Given a multi-view sample $\dot{\mathbf{y}} = \{\mathbf{y}^v\}_{v=1}^V \in \dot{\mathcal{S}}_k$, the representation matrix $\mathbf{C}$ is referred to as a Multi-view Subspace-Preserving Representation of $\dot{\mathbf{y}}$ such that (1) $\mathbf{y}^v = \mathbf{D}^v\mathbf{C}(:,v),\ \forall v \in [V]$, (2) rowsupp$(\mathbf{C}) \subset \mathcal{I}_k$.*

Despite establishing the first theory for MSPR, the work [Wang, 2024] is confined to the equality constrained MJSR model, which may be not suitable for practical use. For instance, the equality constraint $\mathbf{y}^v = \mathbf{D}^v\mathbf{C}(:,v),\ \forall v \in [V]$ is too restricted and hardly to hold even when the data contain noise in only one single view or the sampling in the dictionary $\mathbf{D}^v$ is not sufficient in only one single view such that $\mathbf{y}^v \notin \text{span}(\mathbf{D}^v)$. To broaden the scope of applications of MSPR, in this paper we consider the Regularized Multi-view Joint Sparse Representation (RMJSR) model

$$\min_{\mathbf{C}\in\mathbb{R}^{n\times V}} \|\mathbf{C}\|_{1,2} + \frac{\gamma}{2}\sum_{v=1}^{V} \|\mathbf{y}^v - \mathbf{D}^v\mathbf{C}(:,v)\|_2^2 .$$

Remark 1. *Note that the regularization parameter $\gamma > 0$ should not be too small. Otherwise, the optimal solution will be the zero matrix. So to avoid such trivial solution,*

Notations/Acronyms	Descriptions	Notations/Acronyms	Descriptions
N	number of samples	V	number of views
K	number of classes	d_v	data dimension of the v-th view
$a, \mathbf{a}, \mathbf{A}, \mathcal{A}$	scalar, vector, matrix, set	$\dot{\mathbf{a}} = \{\mathbf{a}^v\}_{v=1}^V, \dot{\mathbf{A}} = \{\mathbf{A}^v\}_{v=1}^V$	multi-view vector, matrix data
$[N]$	set of integers from 1 to N	$\mathbf{A}(i,:)$ and $\mathbf{A}(:,j)$	i-th row and j-th column of matrix $\mathbf{A}$
$\mathbf{D}^v = [\mathbf{d}_1^v, \cdots, \mathbf{d}_N^v] \in \mathbb{R}^{d_v \times N}$	data matrix of the v-th view	$\dot{\mathbf{D}}$	$\{\mathbf{D}^v, \forall v \in [V]\}$
$\mathcal{S}_k^v \subset \mathbb{R}^{d_v}$	the k-th subspace in the v-th view	$\dot{\mathcal{S}}_k$	$\{\mathcal{S}_k^v, \forall v \in [V]\}$
$\mathcal{I}_k \subset [N]$	index set of samples in the k-th class	$\mathcal{I}_{-k} \subset [N]$	index set of samples out of the k-th class
SPR	Subspace-Preserving Representation	MSPR	Multi-view SPR
ITG	Individual Theoretical Guarantee	UTG	Universal Theoretical Guarantee
MSCla	Multi-view Subspace Classification	MSClu	Multi-view Subspace Clustering
$\|\mathbf{a}\|_2 = \sqrt{\sum_i a_i^2}$	ℓ_2 norm of vector $\mathbf{a}$	$\|\dot{\mathbf{a}}\|_{M,2} = \max_v \|\mathbf{a}^v\|_2$	multi-view ℓ_2 norm of $\dot{\mathbf{a}}$
$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_i a_i \cdot b_i$	inner product of vectors $\mathbf{a}$ and $\mathbf{b}$	$\langle \dot{\mathbf{a}}, \dot{\mathbf{b}} \rangle_M = \sum_{v=1}^V \langle \mathbf{a}^v, \mathbf{b}^v \rangle$	multi-view inner product of $\dot{\mathbf{a}}$ and $\dot{\mathbf{b}}$
$\|\mathbf{A}\|_{1,2} = \sum_{i=1}^N \|\mathbf{A}(i,:)\|_2$	$\ell_{1,2}$ norm of $\mathbf{A} \in \mathbb{R}^{N \times V}$	$\|\mathbf{A}\|_{\infty,2} = \max_i \|\mathbf{A}(i,:)\|_2$	$\ell_{\infty,2}$ norm of $\mathbf{A} \in \mathbb{R}^{N \times V}$
$\mathbf{D}_k^v$	sub-matrix of $\mathbf{D}^v$ containing samples from the subspace $\mathcal{S}_k^v$		
$\mathbf{D}_{-k}^v$	sub-matrix of $\mathbf{D}^v$ excluding samples from the subspace $\mathcal{S}_k^v$		
$\dot{\mathbf{D}}_k, \dot{\mathbf{D}}_{-k}$	$\{\mathbf{D}_k^v, \forall v \in [V]\}$ and $\{\mathbf{D}_{-k}^v, \forall v \in [V]\}$		
$\dot{\mathbf{d}}_i^{(k)}$	the i-th multi-view sample in $\dot{\mathbf{D}}_k$		
$\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k$	$\mathbf{y}^v \in \mathcal{S}_k^v$ holds for each view $v \in [V]$		

Table 1: Summary of key notations and acronyms utilized in this paper.

we should set $\gamma \geq t$ for some threshold. In fact, we set $\gamma > 1/\|\dot{\mathbf{D}}^T * \dot{\mathbf{y}}\|_{\infty,2}$ (see Definition 2 for the explanation) in our theory. On the other hand, if $\gamma \to \infty$, the RMJSR model reduces to the MJSR model, which implies that the results in [Wang, 2024] can be seen as special cases of our results in UTG.

To understand the correctness and success of RMJSR for MSCla and MSClu, we raise several interesting questions:

Q1. For a specific *individual* multi-view test sample, what are the theoretical conditions for RMJSR (2) to produce MSPR? We refer to such theoretical guarantee as **Individual Theoretical Guarantee (ITG)**.

Q2. For *all* multi-view test samples, what are the theoretical conditions for RMJSR (2) to yield MSPR? We refer to such theoretical guarantee as **Universal Theoretical Guarantee (UTG)**.

Q3. How to apply the proposed theory for MSCla and MSClu, respectively?

3 Main Results

In this section, we provide positive answers to the questions posed in the last section. To this end, we first introduce some useful concepts, such as multi-view matrix product, multi-view oracle point, multi-view incoherence, etc.

3.1 Necessary Concepts

Definition 2. ***(Multi-view Matrix Product)*** *Consider two multi-view matrix data $\dot{\mathbf{A}} = \{\mathbf{A}^v\}_{v=1}^V$ and $\dot{\mathbf{B}} = \{\mathbf{B}^v\}_{v=1}^V$ where $\mathbf{A}^v \in \mathbb{R}^{m \times d_v}$ and $\mathbf{B}^v \in \mathbb{R}^{d_v \times n}$. The multi-view matrix product of $\dot{\mathbf{A}}$ and $\dot{\mathbf{B}}$ is defined as*

$$\dot{\mathbf{A}} * \dot{\mathbf{B}} = \left[\mathbf{A}^1\mathbf{B}^1, \cdots, \mathbf{A}^V\mathbf{B}^V\right] \in \mathbb{R}^{m \times nV}. \tag{3}$$

Accordingly, when $\dot{\mathbf{B}}$ reduces to a multi-view vector data $\dot{\mathbf{b}} = \{\mathbf{b}^v\}_{v=1}^V$ where $\mathbf{b}^v \in \mathbb{R}^{d_v}$, we have $\dot{\mathbf{A}} * \dot{\mathbf{b}} = \left[\mathbf{A}^1\mathbf{b}^1, \cdots, \mathbf{A}^V\mathbf{b}^V\right] \in \mathbb{R}^{m \times V}$.

Definition 3. ***(Multi-view Oracle Point)*** *For any optimal solution $\mathbf{C}^*$ to the RMJSR model (2), we define the multi-view oracle point $\dot{\mathbf{o}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}, \mathbf{C}^*) = \{\mathbf{o}^{*,v}(\dot{\mathbf{y}}, \dot{\mathbf{D}}, \mathbf{C}^*)\}_{v=1}^V$ such that*

$$\mathbf{o}^{*,v}(\dot{\mathbf{y}}, \dot{\mathbf{D}}, \mathbf{C}^*) := \gamma(\mathbf{y}^v - \mathbf{D}^v\mathbf{C}^*(:,v)), \ \forall v \in [V]. \tag{4}$$

The optimal solutions $\mathbf{C}^*$ to RMJSR and $\dot{\mathbf{u}}^*$ to its dual problem have the following relationship.

Lemma 1. *For each fixed point $\dot{\mathbf{y}}$ and the parameter γ, the dual point $\dot{\mathbf{u}}^* = \dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}})$ is unique and coincides with the multi-view oracle point, i.e., $\dot{\mathbf{u}}^* = \dot{\mathbf{o}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}, \mathbf{C}^*)$.*

Proof. The proof can be seen in the supplementary material due to space limitation. □

Consider the RMJSR model with respect to (w.r.t.) the sub-dictionary $\dot{\mathbf{D}}_k$ as

$$\min_{\mathbf{C} \in \mathbb{R}^{n_k \times V}} \|\mathbf{C}\|_{1,2} + \frac{\gamma}{2} \sum_{v=1}^{V} \|\mathbf{y}^v - \mathbf{D}_k^v \mathbf{C}(:,v)\|_2^2. \tag{5}$$

For ease of presentation, let $\mathbf{C}_k^* = \mathbf{C}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)$ denote any optimal solution to problem (5).

The dual problem of (5) can be formulated as

$$\begin{aligned} &\max_{\dot{\mathbf{u}} = \{\mathbf{u}^v\}_{v=1}^V} \sum_{v=1}^{V} \left(\langle \mathbf{y}^v, \mathbf{u}^v \rangle - \frac{1}{2\gamma} \langle \mathbf{u}^v, \mathbf{u}^v \rangle \right) \\ &\text{s.t.} \ \left\| \left[\left(\mathbf{D}_k^1\right)^T \mathbf{u}^1, \cdots, \left(\mathbf{D}_k^V\right)^T \mathbf{u}^V \right] \right\|_{\infty,2} \leq 1. \end{aligned} \tag{6}$$

Let $\dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)$ denote the optimal solution to problem (6), which is referred to as dual point of the multi-view sample $\dot{\mathbf{y}}$. Note that $\sum_{v=1}^V \left(\langle \mathbf{y}^v, \mathbf{u}^v \rangle - \frac{1}{2\gamma}\langle \mathbf{u}^v, \mathbf{u}^v \rangle\right) = -\sum_{v=1}^V \frac{1}{2\gamma}\|\mathbf{u}^v - \gamma\mathbf{y}^v\|_2^2 + \frac{\gamma}{2}\|\mathbf{y}^v\|_2^2$. The dual problem is

$$\begin{aligned} &\min_{\dot{\mathbf{u}} = \{\mathbf{u}^v\}_{v=1}^V} \sum_{v=1}^{V} \|\mathbf{u}^v - \gamma\mathbf{y}^v\|_2^2, \\ &\text{s.t.} \ \left\| \left[\left(\mathbf{D}_k^1\right)^T \mathbf{u}^1, \cdots, \left(\mathbf{D}_k^V\right)^T \mathbf{u}^V \right] \right\|_{\infty,2} \leq 1. \end{aligned} \tag{7}$$

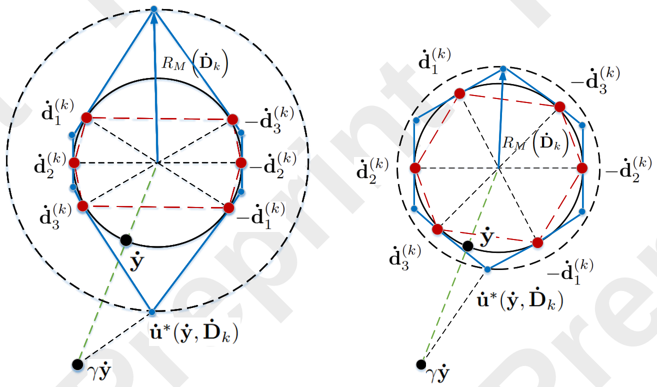

Figure 1: Geometric illustration of $\left\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{M,2}$ and the multi-view circumradius $R_M\left(\dot{\mathbf{D}}_k\right)$ with $\dot{\mathbf{D}}_k=[\dot{\mathbf{d}}_1^{(k)},\dot{\mathbf{d}}_2^{(k)},\dot{\mathbf{d}}_3^{(k)}]$. Left: large $\left\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{M,2}$ and $R_M\left(\dot{\mathbf{D}}_k\right)$; Right: small $\left\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{M,2}$ and $R_M\left(\dot{\mathbf{D}}_k\right)$. Note that they all have small values while the points in $\dot{\mathbf{D}}_k$ are well spread out.

Thus, the optimal solution $\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)$ is the closest point to $\gamma\dot{\mathbf{y}}$ in the dual set $\left\{\dot{\mathbf{u}}:\left\|\dot{\mathbf{D}}_k^T*\dot{\mathbf{u}}\right\|_{\infty,2}\le 1\right\}$.

It can be proved that the primal and dual solutions satisfy

$$\dot{\mathbf{u}}^{*,v}(\dot{\mathbf{y}},\dot{\mathbf{D}}_k):=\gamma(\mathbf{y}^v-\mathbf{D}_k^v\mathbf{C}_k^*(:,v)),\ \forall v\in[V]. \quad (8)$$

If $\dot{\mathbf{y}}\in\dot{\mathcal{S}}_k$, its dual point is also in $\dot{\mathcal{S}}_k$, i.e., $\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\in\dot{\mathcal{S}}_k$.

To measure the correlation among samples from various classes, we give the definition of multi-view incoherence.

Definition 4. *(**Multi-view Incoherence** [Wang, 2024]) The multi-view incoherence between two multi-view datasets $\dot{\mathbf{A}}=\{\mathbf{A}^v\}_{v=1}^V$ and $\dot{\mathbf{B}}=\{\mathbf{B}^v\}_{v=1}^V$ is defined as*

$$\mu_M\left(\dot{\mathbf{A}},\dot{\mathbf{B}}\right)=\max_{\dot{\mathbf{a}}\in\dot{\mathbf{A}},\dot{\mathbf{b}}\in\dot{\mathbf{B}}}\left\|\left[\frac{\langle\mathbf{a}^1,\mathbf{b}^1\rangle}{\|\mathbf{a}^1\|_2\|\mathbf{b}^1\|_2},\cdots,\frac{\langle\mathbf{a}^V,\mathbf{b}^V\rangle}{\|\mathbf{a}^V\|_2\|\mathbf{b}^V\|_2}\right]\right\|_2, \quad (9)$$

where $\dot{\mathbf{a}}=\{\mathbf{a}^v\}_{v=1}^V$ and $\dot{\mathbf{b}}=\{\mathbf{b}^v\}_{v=1}^V$.

To characterize the distribution of the samples of $\dot{\mathbf{D}}_k$ in $\dot{\mathcal{S}}_k$, we also introduce the multi-view circumradius $R_M\left(\dot{\mathbf{D}}_k\right)$.

Definition 5. *(**Multi-view Circumradius** [Wang, 2024]) The multi-view circumradius of the multi-view sub-dictionary $\dot{\mathbf{D}}_k=\{\mathbf{D}_k^v\}_{v=1}^V$ is defined as*

$$R_M\left(\dot{\mathbf{D}}_k\right)=\max_{\dot{\mathbf{u}}\in\dot{\mathcal{S}}_k}\|\dot{\mathbf{u}}\|_{M,2},\ \text{s.t.}\ \left\|\dot{\mathbf{D}}_k^T*\dot{\mathbf{u}}\right\|_{\infty,2}\le 1. \quad (10)$$

The multi-view circumradius $R_M\left(\dot{\mathbf{D}}_k\right)$ characterizes the distribution of the samples of $\dot{\mathbf{D}}_k$ in the corresponding multi-view subspace $\dot{\mathcal{S}}_k$.

3.2 Individual Theoretical Guarantee

With the concepts above, we are now ready to give the first answer to the question (Q1) on the ITG for multi-view subspace-preserving recovery.

Theorem 1. *The solution to the RMJSR model (2) is a MSPR for any fixed $\dot{\mathbf{y}}\in\dot{\mathcal{S}}_k\backslash\{\dot{\mathbf{0}}\}$, $k\in[K]$ if*

$$\mu_M\left(\dot{\mathbf{D}}_{-k},\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right)\left\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{M,2}<1, \quad (11)$$

where $\mathbf{C}_k^*=\mathbf{C}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)$ is the solution to the model (5), and

$$\dot{\mathbf{u}}^{*,v}(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)=\gamma(\mathbf{y}^v-\mathbf{D}_k^v\mathbf{C}_k^*(:,v)),\ \forall v\in[V]. \quad (12)$$

Proof. The proof can be seen in the supplementary material due to space limitation. □

Remark 2. *Theorem 1 reveals important geometric insights for the successful condition of RMJSR to achieve MSPR. Recall that the inter-class multi-view coherence $\mu_M\left(\dot{\mathbf{D}}_{-k},\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right)$ is small if the samples in $\dot{\mathbf{D}}_{-k}$ are away from $\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)$, which is in $\dot{\mathcal{S}}_k$ if $\dot{\mathbf{y}}\in\dot{\mathcal{S}}_k$. $\left\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{M,2}$ is small if the samples in $\dot{\mathbf{D}}_k$ are well spread out. Therefore, Theorem 1 ensures that the solution to RMJSR is a MSPR if the samples in other classes are not too close to the dual point $\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)$ of the sample $\dot{\mathbf{y}}$ and meanwhile the samples in $\dot{\mathbf{D}}_k$ from the same class of $\dot{\mathbf{y}}$ are well spread out. We refer to the condition (11) as the Individual Multi-view Recovery Condition (IMRC).*

To provide more intuitive geometric insights, we present another answer to the question (Q1).

Theorem 2. *The solution to the RMJSR model (2) is a MSPR for any fixed $\dot{\mathbf{y}}\in\dot{\mathcal{S}}_k\backslash\{\dot{\mathbf{0}}\}$, $k\in[K]$ if*

$$\left\|\dot{\mathbf{D}}_{-k}^T*\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{\infty,2}<\left\|\dot{\mathbf{D}}_k^T*\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{\infty,2}, \quad (13)$$

where $\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)$ is defined in Theorem 1.

Proof. The proof can be seen in the supplementary material due to space limitation. □

Remark 3. *Theorem 2 asserts that the multi-view subspace-preserving recovery of RMJSR is guaranteed to succeed if the multi-view dual point $\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)$ of the test sample $\dot{\mathbf{y}}$ is closer to the sub-dictionary $\dot{\mathbf{D}}_k$ in the same class of $\dot{\mathbf{y}}$ than the sub-dictionary $\dot{\mathbf{D}}_{-k}$ in other classes. For simplicity, we refer to $\left\|\dot{\mathbf{D}}_{-k}^T*\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{\infty,2}$ and $\left\|\dot{\mathbf{D}}_k^T*\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\right\|_{\infty,2}$ as the mismatched multi-view coherence and the matched multi-view coherence, respectively. Thus, the multi-view coherence $\|\cdot\|_{\infty,2}$ plays a role as a similarity metric. Compared with Theorem 1, the theoretical condition in Theorem 2 is more concise and geometrically intuitive.*

Geometric Interpretation of ITG. The concise geometric characterizations can help us better understand the theoretical results of ITG. As shown in Fig. 1, $\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}},\dot{\mathbf{D}}_k)\|_{M,2}$ is small when the samples in $\dot{\mathbf{D}}_k$ are well spread out.

On the other hand, the inter-class multi-view coherence $\mu_M(\dot{\mathbf{D}}_{-k}, \dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k))$ is small if the samples in $\dot{\mathbf{D}}_{-k}$ are away from $\dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)$. This geometrically characterizes when the condition (11) in Theorem 1 satisfies.

Moreover, Theorem 2 has a more concise and intuitive geometrical characterization. As shown in Fig. 2, the condition (13) holds if and only if $\theta_k < \theta_{-k} = \min\{\alpha_1, \alpha_2\}$ (α_1, α_2 denote the included angles of $\dot{\mathbf{u}}^*$ and the subspace $\dot{\mathcal{S}}_1$ and $\dot{\mathcal{S}}_2$, respectively). This illustrates that the multi-view dual point $\dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)$ of the test sample $\dot{\mathbf{y}}$ should be closer to the sub-dictionary $\dot{\mathbf{D}}_k$ in the same class of $\dot{\mathbf{y}}$ than the sub-dictionary $\dot{\mathbf{D}}_{-k}$ in other classes to guarantee the MSPR.

3.3 Universal Theoretical Guarantee

The first Universal Theoretical Guarantee (UTG) of RMJSR for MSPR is obtained by leveraging on Theorem 1 and deriving the upper bounds of $\mu_M\left(\dot{\mathbf{D}}_{-k}, \dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)\right)$ and $\left\|\dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)\right\|_{M,2}$, respectively. Let $\dot{\mathcal{U}}_k$ denote the set of the dual points of data samples in $\dot{\mathcal{S}}_k$, i.e., $\dot{\mathcal{U}}_k := \{\dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k) | \forall \dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}\}$.

With the definitions above, we have the following result.

Theorem 3. *The solution to (2) is a MSPR for all $k \in [K]$ and any multi-view sample $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$ if*

$$\mu_M\left(\dot{\mathbf{D}}_{-k}, \dot{\mathcal{U}}_k\right) R_M\left(\dot{\mathbf{D}}_k\right) < 1, \; k = 1, \cdots, K. \tag{14}$$

Proof. The proof can be seen in the supplementary material due to space limitation. □

Note from Eq. (8) that $\dot{\mathbf{u}}^{*,v}(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k) \in \dot{\mathcal{S}}_k$ for $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k$ and $\dot{\mathbf{D}}_k \subset \dot{\mathcal{S}}_k$. Thus, $\dot{\mathcal{U}}_k \subset \dot{\mathcal{S}}_k$. Then we have the next corollary.

Corollary 1. *The solution to (2) is a MSPR for all $k \in [K]$ and any multi-view sample $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$ if*

$$\mu_M\left(\dot{\mathbf{D}}_{-k}, \dot{\mathcal{S}}_k\right) R_M\left(\dot{\mathbf{D}}_k\right) < 1, \; k = 1, \cdots, K. \tag{15}$$

Remark 4. *The results in [Wang, 2024] can be seen as a special case of Theorem 3 and Corollary 1 when the parameter $\gamma \to \infty$. When γ is large enough, the optimal solution of (2) is also the optimal solution of the problem (1). For simplicity, we refer to the condition (15) as the Universal Multi-view Recovery Condition (UMRC).*

Geometric Interpretation of UTG. The theoretical results of UTG can also be geometrically characterized in a concise manner. As shown in Fig. 1, the multi-view circumradius $R_M(\dot{\mathbf{D}}_k)$ is small when the samples in $\dot{\mathbf{D}}_k = \left[\dot{\mathbf{d}}_1^{(k)}, \dot{\mathbf{d}}_2^{(k)}, \dot{\mathbf{d}}_3^{(k)}\right]$ are well spread out. Besides, when the samples in $\dot{\mathbf{D}}_{-k}$ are away from the subspace $\dot{\mathcal{S}}_k$, the class coherence $\mu_M(\dot{\mathbf{D}}_{-k}, \dot{\mathcal{S}}_k)$ is small. Thus, this geometric interpretation implies the condition (15) in Corollary 1 under which the solution of RMJSR is a MSPR. The samples in other classes $\dot{\mathbf{D}}_{-k}$ should be not too close to the subspace $\dot{\mathcal{S}}_k$ which the sample $\dot{\mathbf{y}}$ is drawn from. Meanwhile, the samples in $\dot{\mathbf{D}}_k$ from the same class of $\dot{\mathbf{y}}$ should be well spread out. These geometrical insights provide a better comprehension for the theoretical results in UTG.

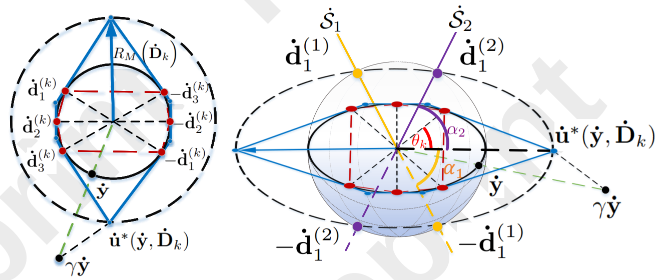

Figure 2: Geometric interpretations of subspace-preserving conditions in Theorem 1 and 2. Left: the samples $\dot{\mathbf{d}}_1^{(k)}, \dot{\mathbf{d}}_2^{(k)}, \dot{\mathbf{d}}_3^{(k)}$ from $\dot{\mathbf{D}}_k$ lying on the unit circle of a 2-dimensional subspace $\dot{\mathcal{S}}_0$ and the dual point $\dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)$ of a fixed $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_0 \backslash \{\dot{\mathbf{0}}\}$; Right: small $\mu_M\left(\dot{\mathbf{D}}_{-k}, \dot{\mathbf{u}}^*(\dot{\mathbf{y}}, \dot{\mathbf{D}}_k)\right)$ while the samples $\dot{\mathbf{d}}_1^{(1)}, \dot{\mathbf{d}}_1^{(2)}$ are away from the dual point $\dot{\mathbf{u}}^*$, which come from other subspaces (1-dimensional $\dot{\mathcal{S}}_1$ and $\dot{\mathcal{S}}_2$) in $\dot{\mathbf{D}}_{-k}$ lying on the unit sphere in the ambient space $\mathbb{R}^3$. Another concise interpretation is that the angle $\theta_k < \theta_{-k} = \min\{\alpha_1, \alpha_2\}$ when the samples $\dot{\mathbf{d}}_1^{(1)}, \dot{\mathbf{d}}_1^{(2)}$ are away from the dual point $\dot{\mathbf{u}}^*$. Then the Eq. (15) holds and the solution to RMJSR is a MSPR for any fixed $\dot{\mathbf{y}}$.

3.4 Discussion and Comparison with Other Works

To further highlight the novelty of this work, we provide an in-depth discussion on the relationship between our results with other works on SPR. Note that most prior works require that the **equality constraint** strictly satisfies, and they only provide the **UTG**. Tab. 2 summarizes the main differences between prior works and this work.

Comparison to [Kaba *et al.*, 2021]. [Kaba *et al.*, 2021] derives another condition for SPR, termed as subspace nullspace property (SNSP), which is defined as follows.

Definition 6. *Given $\mathcal{A}$ as a partition of the index set $\{1, \cdots, N\}$ and $\mathrm{Null}(\mathbf{D})$ as the nullspace of $\mathbf{D}$, a subset of $\mathrm{Null}(\mathbf{D})$ w.r.t. (with respect to) $\mathcal{A}$ is denoted as*

$$\mathrm{Null}(\mathbf{D}, \mathcal{A}) := \{\boldsymbol{\xi} \in \mathrm{Null}(\mathbf{D}) : \mathrm{supp}(\mathbf{D}) \not\subseteq P, \; \forall A \in \mathcal{A}\}.$$

Let A^c be the complement set of A. For any $\boldsymbol{\xi} \in \mathrm{Null}(\mathbf{D}, \mathcal{A})$ and $A \in \mathcal{A}$, the dictionary matrix $\mathbf{D}$ satisfies SNSP such that

$$\min_{\mathbf{c}:\mathbf{D}_A\mathbf{c} = \mathbf{D}_A\boldsymbol{\xi}_A} \|\mathbf{z}\|_1 < \|\boldsymbol{\xi}_{A^c}\|_1. \tag{16}$$

Compared with SNSP, our conditions have concise and intuitive geometric characterizations with both UTG and ITG. Besides, their result is limited to single-view data while our results can be directly used for multi-view data. Finally, their result holds only if the equality constraint satisfies while our results can hold without such strong assumption.

Comparison to [Thaker *et al.*, 2022]. In [Thaker *et al.*, 2022], the authors consider the application of block sparse recovery model for the reverse engineering adversarial attacks. There are several key differences between our work and the work [Thaker *et al.*, 2022]. *Firstly*, our theory is devised for multi-view data while [Thaker *et al.*, 2022] is for single-view data, which can not be directly applied for multi-view data. *Secondly*, the main results in Theorem 1 and 3 in our paper both include the UTG and ITG for multi-view subspace-preserving recovery while the core results in

Theory	References	Year	Model	UTG	ITG	w/o equality constraint
Single-view SPR	[Soltanolkotabi and Candés, 2012]	2012	BP	✓	×	×
	[Dyer *et al.*, 2013]	2013	OMP	✓	×	×
	[You and Vidal, 2015]	2015	BP&OMP	✓	×	×
	[You *et al.*, 2016]	2016	OMP	✓	×	×
	[Wang *et al.*, 2019]	2019	AR	✓	×	×
	[Kaba *et al.*, 2021]	2021	BP	✓	×	×
	[Thaker *et al.*, 2022]	2022	BSP	✓	×	×
Multi-view SPR	[Wang, 2024]	2024	MJSR	✓	×	×
	This paper	2025	RMJSR	✓	✓	✓

Table 2: Comparison of theoretical works of Subspace-Preserving Recovery (SPR). Here "UTG" and "ITG" mean whether the works provide universal and individual theoretical guarantee for SPR, respectively. "w/o equality constraint" means whether the results hold without the equality constraint, which is too strict in practical use. Specifically, conditions and results of UTG are for all samples in corresponding subspaces. Instead, ITG individually devises conditions and guarantee for a fixed data sample, which is more flexible in practical use.

[Thaker *et al.*, 2022] only consider the UTG for single-view subspace-preserving recovery. *Thirdly*, the results in [Thaker *et al.*, 2022] requires both the label information of data in anticipation and the equality constraint, while our results do not depend on these conditions.

More representative works of SPR theory can be found in [You and Vidal, 2015; Soltanolkotabi and Candés, 2012; You *et al.*, 2016; Robinson *et al.*, 2019; Wang *et al.*, 2019; You *et al.*, 2019]. Although advancing our understanding, all these works are confined to **single-view data** and cannot be directly applied to multi-view data. Besides, their results hold only if the equality constraint strictly satisfies while our results do not rely on this strong assumption.

Comparison to [Wang, 2024]. The work [Wang, 2024] devotes to extending the SPR theory to **multi-view data** and establishing the Atomic Recovery Property (ARP) under which the multi-view joint sparse representation model can obtain a multi-view subspace-preserving representation.

To our best knowledge, [Wang, 2024] is the only existing theoretical guarantee for multi-view SPR. However, its results are confined to the equality constrained MJSR model, which may be not suitable for practical use. For instance, the equality constraint $\mathbf{y}^v = \mathbf{D}^v\mathbf{C}(:,v),\ \forall v \in [V]$ is too restricted and hardly to hold even when the data contain noise in only one single view or the sampling in the dictionary $\mathbf{D}^v$ is not sufficient in only one single view such that $\mathbf{y}^v \notin \text{span}(\mathbf{D}^v)$. Compared with [Wang, 2024], our theory does not rely on this strong assumption and is more flexible in real-world applications. Moreover, [Wang, 2024] only provides the universal theoretical guarantee (UTG). As demonstrated before, the analysis of UTG requires all samples in the sub-matrices $\mathbf{D}_k^v$ and $\mathbf{D}_{-k}^v, v \in [V]$ to satisfy the UMRC condition. This is too strict, especially for one fixed multi-view sample $\dot{\mathbf{y}}$. Thus, we provide both UTG and ITG in this paper.

4 Application to Multi-view Subspace Classification and Clustering

For any new multi-view test sample $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$, Regularized Multi-view Subspace Classification (RMSCla) method, a variant of [Shekhar *et al.*, 2013], calculates the representation matrix of $\dot{\mathbf{y}}$ by solving the following model

$$\min_{\mathbf{C}\in\mathbb{R}^{N\times V}} \|\mathbf{C}\|_{1,2} + \frac{\gamma}{2}\sum_{v=1}^{V} \|\mathbf{y}^v - \mathbf{D}^v\mathbf{C}(:,v)\|_2^2,\ \forall v \in [V],$$

where $\dot{\mathbf{D}} = \{\mathbf{D}^v\}_{v=1}^V$ denotes the multi-view training data. The reconstruction residual of each class is calculated as

$$r_k(\dot{\mathbf{y}}) = \sum_{v=1}^{V} \|\mathbf{y}^v - \mathbf{D}^v\delta_k(\mathbf{C}(:,v))\|_2,\ k = 1,\cdots,K \quad (17)$$

where $\delta_k(\mathbf{C}(:,v)) \in \mathbb{R}^N$ denotes the vector containing the entries associated with the k-th class and changing the remaining entries as zeros.

Theorem 4. *The RMSCla method is guaranteed to succeed to classify any new multi-view test sample $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k\backslash\{\dot{\mathbf{0}}\}$ for any class k if the multi-view training data $\dot{\mathbf{D}}$ satisfies the UMRC.*

Proof. The proof is in the supplementary material. □

Consider the Regularized Multi-view Subspace Clustering (RMSClu) model, which is a variant of Multi-view Sparse Subspace Clustering in [Wang *et al.*, 2023]

$$\min_{\mathbf{C}_i\in\mathbb{R}^{N\times V}} \|\mathbf{C}_i\|_{1,2} + \frac{\gamma}{2}\sum_{v=1}^{V} \|\mathbf{y}_i^v - \mathbf{Y}^v\mathbf{C}_i(:,v)\|_2^2, \quad \text{s.t. } \mathbf{C}_i(i,:) = \mathbf{0},\ v \in [V], \quad (18)$$

where $\dot{\mathbf{Y}} = \{\mathbf{Y}^v\}_{v=1}^V, \mathbf{Y}^v = (\mathbf{y}_1^v,\cdots,\mathbf{y}_N^v)$ denotes the multi-view data which requires clustering. Let $\mathbf{C}_i^*$ be the optimal solution of (18) for $i \in [N]$. We first construct the common representation matrix $\mathbf{C}^* \in \mathbb{R}^{N\times N}$ such that

$$\mathbf{C}^*(:,i) = [\|\mathbf{C}_i^*(1,:)\|_2,\cdots,\|\mathbf{C}_i^*(N,:)\|_2]^T, \quad (19)$$

for $i \in [N]$. Then we compute the similarity matrix $\mathbf{Z} = (|\mathbf{C}^*| + |\mathbf{C}^{*T}|)/2$. Finally, we apply the spectral clustering algorithm to the similarity matrix and obtain the final clustering results. A desirable similarity matrix $\mathbf{Z}$ should satisfy the following subspace-preserving property: $Z_{ij} \neq 0$ only if the i-th sample and the j-th sample are from the same cluster.

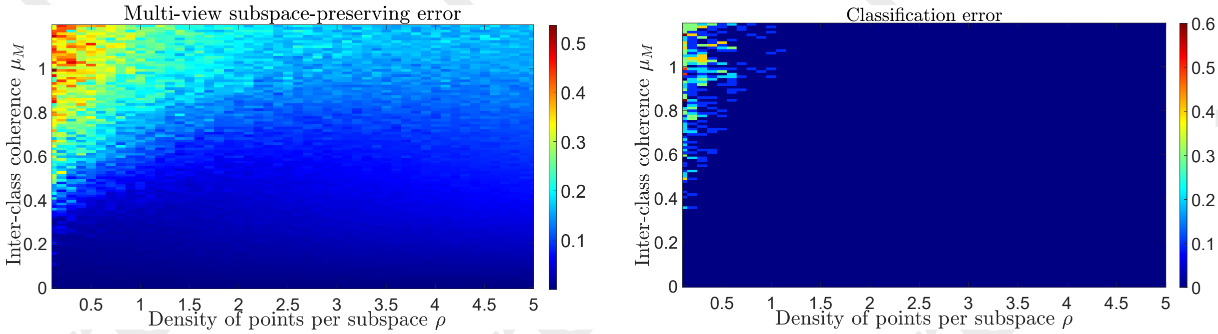

Figure 3: Error metrics under different values of the inter-class coherence μ_M and sampling density ρ for data points in each subspace. Left: multi-view subspace-preserving error; Right: classification error.

Theorem 5. *The similarity matrix* $\mathbf{Z}$ *generated by the RMSClu model (18) satisfies the subspace-preserving property if and only if the leave-one-out multi-view data* $\dot{\mathbf{Y}}\backslash\dot{\mathbf{y}}_i$ *satisfies the UMRC for* $i \in [N]$.

Proof. The proof can be seen in the supplementary material due to space limitation. □

5 Numerical Experiments

5.1 Evaluation Metrics

We consider two metrics including the multi-view subspace-preserving (MSP) error and the classification error. Given any multi-view test sample $\dot{\mathbf{y}}$, we denote the representation matrix obtained by RMJSR as $\mathbf{C}^*$. If $\dot{\mathbf{y}}$ is drawn from the class k, then the MSP error is calculated as $\text{MSP error} = 1 - \|\delta_k(\mathbf{C}^*)\|_{1,2}/\|\mathbf{C}^*\|_{1,2}$, where $\delta_k(\mathbf{C}^*)$ denotes the matrix which only keeps the rows of $\mathbf{C}^*$ aligned with the k-th class and other rows are all set to zero vectors. Hence, the MSP error is utilized to measure how far the matrix $\mathbf{C}^*$ from being a multi-view subspace-preserving representation (MSPR). The MSP error is smaller while $\mathbf{C}^*$ is closer to a MSPR. Another metric is the classification error, which is calculated as the percent of misclassified test samples.

5.2 Impact of Inter-class Coherence and Sampling Density for MSCla

In the aforementioned analysis, we demonstrate that the success of RMJSR for multi-view subspace classification relies on the inter-class coherence and the distribution of data points in each subspace. To validate the argument above, we generate three subspaces with two views (the number of views $V = 2$) $\{\mathcal{S}_k^v\}_{k=1}^3, v \in [2]$ such that the subspaces in each view share the same dimension. Then we set the subspace dimension $d_1 = 20, d_2 = 40$ while the ambient dimension $m_1 = 40, m_2 = 80$, i.e., $m_v = 2d_v$. More results of the cases $m_v < 2d_v$ or $m_v > 2d_v$ and $V = 3$ are shown in the supplementary material. Following [Wang *et al.*, 2019], the subspace bases $\mathbf{U}_k^v \in \mathbb{R}^{m_v \times d_v}, v \in [2]$ are generated as

$$\mathbf{U}_1^v = \begin{bmatrix} \mathbf{I}_{d_v} \\ \mathbf{0}_{d_v \times d_v} \end{bmatrix}, \quad \mathbf{U}_2^v = \begin{bmatrix} \mathbf{0}_{d_v \times d_v} \\ \mathbf{I}_{d_v} \end{bmatrix},$$

$$\mathbf{U}_3^v = \begin{bmatrix} \cos(\varphi_1) & 0 & \cdots & 0 \\ 0 & \cos(\varphi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cos(\varphi_{d_v}) \\ \sin(\varphi_1) & 0 & \cdots & 0 \\ 0 & \sin(\varphi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sin(\varphi_{d_v}) \end{bmatrix},$$

where $\mathbf{I}_{d_v}$ and $\mathbf{0}_{d_v \times d_v}$ denote the identity matrix and zero matrix of size $d_v \times d_v$, respectively. The angles $\{\varphi_i\}_{i=1}^{d_v}$ are set as $\cos\varphi_i = \left(1 - \frac{i-1}{2(d_v-1)}\cos\varphi\right)$ and φ is a tuning parameter. Here, we refer to the inter-class coherence $\mu_M(\dot{\mathbf{D}}_{-k}, \dot{\mathcal{S}}_k)$ as μ_M for simplicity. To study the impact of μ_M, φ varies from 0 to $\pi/2$ and we then calculate the corresponding value of μ_M. Here, ρ denotes the sampling density in each subspace and we randomly sample ρd_1 points from each subspace to construct the training set $\dot{\mathbf{D}}$. The test set contains $N = 500$ samples randomly drawn from each subspace. To study the impact of sampling density, we vary ρ from 0.1 to 5.

Fig. 3 illustrates the average MSP error and the classification error of the RMSCla (RMJSR based MSCla) with various values of μ_M and ρ. The experimental result is consistent with our theoretical analysis, which demonstrates that it is harder for RMJSR to achieve MSPR and accurate classification results when the class coherence is larger and the sampling density of training data in each subspace declines.

6 Conclusions

This paper focuses on the multi-view subspace-preserving recovery theory, which is the theoretical underpinnings for multi-view subspace classification (MSCla) and multi-view subspace clustering (MSClu). Specifically, we prove that the optimal solution to the RMJSR model is a Multi-view Subspace-Preserving Representation (MSPR) under proper conditions. Unlike previous results, our results do not require the equality constraint, which is a common requirement and may be too restrictive in reality. Besides, we provide both universal theoretical guarantee and individual theoretical guarantee for MSPR. The results provide concise geometric characterizations for the success of MSCla and MSClu.

Ethical Statement

There are no ethical issues.

Acknowledgments

We are grateful to the anonymous IJCAI reviewers for their constructive comments. This work was supported by the National Natural Science Foundation of China (under Grant Nos. 62276111, 62076041, 62302188, 72471008, 72434005, and 62202009), the University of Macau (under Grant Nos. MYRG-GRG2024-00290-FST-UMDF and MYRG-CRG2024-00046-FHS), and the HZAU-AGIS Cooperation Fund (under Grant No. SZYJY2023010).

References

[Basri and Jacobs, 2003] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[Chang *et al.*, 2024] Wei Chang, Huimin Chen, Feiping Nie, Rong Wang, and Xuelong Li. Tensorized and compressed multi-view subspace clustering via structured constraint. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):10434–10451, 2024.

[Costeira and Kanade, 1998] João Paulo Costeira and Takeo Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29:159–179, 1998.

[Dyer *et al.*, 2013] Eva L Dyer, Aswin C Sankaranarayanan, and Richard G Baraniuk. Greedy feature selection for subspace clustering. *The Journal of Machine Learning Research*, 14(1):2487–2517, 2013.

[Elhamifar and Vidal, 2009] Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2790–2797, 2009.

[Elhamifar and Vidal, 2010] Ehsan Elhamifar and René Vidal. Clustering disjoint subspaces via sparse representation. In *Proc. IEEE Conf. ICASSP*, pages 1926–1929, 2010.

[Elhamifar and Vidal, 2013] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(11):2765–2781, Nov. 2013.

[Kaba *et al.*, 2021] Mustafa D Kaba, Chong You, Daniel P Robinson, Enrique Mallada, and Rene Vidal. A nullspace property for subspace-preserving recovery. In *International Conference on Machine Learning*, pages 5180–5188, 2021.

[Li *et al.*, 2017] Chun-Guang Li, Chong You, and René Vidal. Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework. *IEEE Transactions on Image Processing*, 26(6):2988–3001, 2017.

[Ma *et al.*, 2008] Yi Ma, Allen Y. Yang, Harm Derksen, and Robert Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3):413–458, 2008.

[Robinson *et al.*, 2019] Daniel P Robinson, Rene Vidal, and Chong You. Basis pursuit and orthogonal matching pursuit for subspace-preserving recovery: Theoretical analysis. *arXiv preprint arXiv:1912.13091*, 2019.

[Shekhar *et al.*, 2013] Sumit Shekhar, Vishal M Patel, Nasser M Nasrabadi, and Rama Chellappa. Joint sparse representation for robust multimodal biometrics recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 36(1):113–126, 2013.

[Soltanolkotabi and Candés, 2012] Mahdi Soltanolkotabi and Emmanuel J Candés. A geometric analysis of subspace clustering with outliers. *The Annals of Statistics*, 40(4):2195–2238, 2012.

[Thaker *et al.*, 2022] Darshan Thaker, Paris Giampouras, and René Vidal. Reverse engineering ℓ_p attacks: A block-sparse optimization approach with recovery guarantees. In *International Conference on Machine Learning*, pages 21253–21271, 2022.

[Vidal, 2011] René Vidal. Subspace clustering. *IEEE Signal Process. Mag.*, 28(2):52–68, Mar. 2011.

[Wang and Xu, 2016] Y. Wang and H. Xu. Noisy sparse subspace clustering. *J. Mach. Learn. Res.*, 17(12):1–41, 2016.

[Wang *et al.*, 2019] Yulong Wang, Yuan Yan Tang, Luoqing Li, Hong Chen, and Jianjia Pan. Atomic representation-based classification: theory, algorithm, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):6–19, 2019.

[Wang *et al.*, 2023] Yulong Wang, Kit Ian Kou, Hong Chen, Yuan Yan Tang, and Luoqing Li. Simultaneous robust matching pursuit for multi-view learning. *Pattern Recognition*, 134:109100, 2023.

[Wang, 2024] Yulong Wang. Atomic recovery property for multi-view subspace-preserving recovery. In *IJCAI*, pages 5144–5152, 2024.

[Wright *et al.*, 2009] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(2):210–227, Jan. 2009.

[Yang *et al.*, 2024] Yang Yang, Yongqiang Tang, Jiangbo Bai, Lu Zhang, and Wensheng Zhang. Multiview subspace tensor self-representation for sar image semi-supervised classification. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[You and Vidal, 2015] Chong You and René Vidal. Geometric conditions for subspace-sparse recovery. In *International conference on machine learning*, pages 1585–1593. PMLR, 2015.

[You *et al.*, 2016] Chong You, Daniel Robinson, and René Vidal. Scalable sparse subspace clustering by orthogonal matching pursuit. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3918–3927, 2016.

[You *et al.*, 2019] Chong You, Chun-Guang Li, Daniel Robinson, and Rene Vidal. Is an affine constraint needed for affine subspace clustering? In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9914–9923, 2019.

[Zhang *et al.*, 2012] Amy Zhang, Nadia Fawaz, Stratis Ioannidis, and Andrea Montanari. Guess who rated this movie: identifying users through subspace clustering. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, page 944–953, 2012.

[Zhang *et al.*, 2020] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. Generalized latent multi-view subspace clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(1):86–99, 2020.

[Zhang *et al.*, 2022] Chao Zhang, Huaxiong Li, Chunlin Chen, Yuhua Qian, and Xianzhong Zhou. Enhanced group sparse regularized nonconvex regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(5):2438–2452, 2022.

[Zhang *et al.*, 2024] Hengmin Zhang, Jian Yang, Jianjun Qian, Guangwei Gao, Xiangyuan Lan, Zhiyuan Zha, and Bihan Wen. Efficient image classification via structured low-rank matrix factorization regression. *IEEE Transactions on Information Forensics and Security*, 19:1496–1509, 2024.