Atomic Recovery Property for Multi-view Subspace-Preserving Recovery

Yulong Wang[1,2,3]
[1] College of Informatics, Huazhong Agricultural University, China
[2] Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education, China
[3] Key Laboratory of Smart Farming Technology for Agricultural Animals, Ministry of Agriculture and Rural Affairs, China
wangyulong6251@gmail.com

Abstract

As the theoretical underpinnings for subspace clustering and classification, subspace-preserving recovery has attracted intensive attention in recent years. However, previous theoretical advances for subspace-preserving recovery only focus on the single-view data and most of them are based on conditions that are only sufficient. In this paper, we propose a necessary and sufficient condition referred to as Atomic Recovery Property (ARP) for multi-view subspace-preserving recovery. To this end, we generalize the atomic norm from single-view data to multi-view data and define the Multi-view Atomic Norm (MAN). Our another contribution is to provide a geometrically more interpretable characterization of ARP with respect to the unit ball of MAN. Based on the proposed multi-view subspace-preserving recovery theory, we also derive novel theoretical results for multi-view subspace clustering and classification, respectively.

1 Introduction

In many machine learning problems, data in a class lie in a low-dimensional subspace of the high-dimensional ambient space [Li *et al.*, 2019; Wright *et al.*, 2009; Elhamifar and Vidal, 2013; Vidal, 2011]. The examples include but not limited to trajectories of a rigidly moving object in a video and facial images of a subject captured under varying illumination. Accordingly, a data set from multiple classes lie in a union of low-dimensional subspaces, where each subspace corresponds to one class. To exploit the low-dimensional structure of high-dimensional data, in recent years a number of supervised and unsupervised learning methods have been developed for subspace classification [Wright *et al.*, 2009] and subspace clustering [Vidal, 2011], respectively.

1.1 Single-view Subspace-Preserving Recovery

As the theoretical underpinnings for subspace clustering and classification, subspace-preserving recovery has attracted intensive attention in recent years [You and Vidal, 2015; Kaba *et al.*, 2021; Zhang *et al.*, 2021]. Let $\mathbf{A} \in \mathbb{R}^{d\times N}$ be a matrix with the columns drawn from multiple low-dimensional subspaces and $\mathbf{y} \in \mathbb{R}^d$ be a data point from one of these subspaces. Subspace-preserving recovery is to find a representation of $\mathbf{y}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$ and the nonzero entries of $\mathbf{x}$ correspond to the columns of $\mathbf{A}$ lying in the same subspace as $\mathbf{y}$. Such a s representation $\mathbf{x}$ is dubbed subspace-preserving [You and Vidal, 2015].

Many pioneering works [Elhamifar and Vidal, 2009; Elhamifar and Vidal, 2010; Soltanolkotabi and Candés, 2012] study the subspace-preserving recovery problem in the context of subspace clustering. In [Elhamifar and Vidal, 2009; Elhamifar and Vidal, 2010], the authors showed that Basis Pursuit (BP) algorithm is guaranteed to yield Subspace-preserving representation (SPR) when the data subspaces are independent and disjoint, respectively. The work [Soltanolkotabi and Candés, 2012] extends the theoretical condition to the case when the subspaces of interest intersect. In [You and Vidal, 2015], the authors identified two sufficient geometric conditions for both BP and OMP (Orthogonal Matching Pursuit) to achieve subspace-preserving recovery. Inspired by the nullspace property [Foucart and Rauhut, 2013] in classical sparse recovery, Kaba *et al.* [Kaba *et al.*, 2021] derived the subspace nullspace property for subspace-preserving recovery. To recover the clean signal from its attacked version and determine the attack type, Thaker *et al.* [Thaker *et al.*, 2022] posed the problem as a block-sparse recovery (BSR) and derived geometric conditions for the success of BSR for the problem. More related works can be found in [Dyer *et al.*, 2013; You *et al.*, 2016; Wang and Xu, 2016; Tsakiris and Vidal, 2018; Wang *et al.*, 2019].

Although advancing our understanding, these prior works on the subspace-preserving recovery theory are only confined to single-view data and cannot be applied for multi-view data, which are widespread in various applications such as multimodal biometrics recognition [Shekhar *et al.*, 2013; Ren *et al.*, 2022] and multi-view clustering [Abavisani and Patel, 2018; Tang and Liu, 2022; Guo *et al.*, 2022; Chen *et al.*, 2023].

1.2 Multi-view Subspace Classification and Clustering

In many real-world applications, data is usually multi-view in nature, such as a video containing both audio signals and visual frames and images of a 3D object captured in distinct viewpoints. To exploit the consistent and complementary information among multiple views, various multi-view subspace clustering and multi-view subspace classification methods have been developed in recent years [Shekhar *et al.*, 2013; Wang *et al.*, 2023]. For instance, Shekhar *et al.* [Shekhar *et al.*, 2013] developed a Joint Sparse Representation (JSR) based multi-view subspace classification method for multimodal biometrics recognition. Wang *et al.* [Wang *et al.*, 2023] proposed a multi-view sparse subspace clustering (MSSC) method based on JSR. More related works can be found in the references [Bahrampour *et al.*, 2015; Zhang *et al.*, 2017; Abavisani and Patel, 2018].

Despite their empirical success and superiority over single-view methods, few theoretical results are reported to justify their effectiveness. In this work, we introduce a new notion dubbed Multi-view Subspace-Preserving Recovery, which plays a central role in the success of multi-view subspace clustering and classification methods. However, it is unclear under what conditions these multi-view subspace clustering and classification methods generate Multi-view Subspace-Preserving Representation (MSPR). The goal of this work is to fill this gap.

1.3 Paper Contributions

In this paper, we make the following contributions:

1. We derive a necessary and sufficient condition termed as Atomic Recovery Property (ARP for short) for multi-view subspace-preserving recovery. To the best of our knowledge, this is the first work to address the multi-view subspace-preserving recovery problem.
2. We provide a geometrically more interpretable characterization of ARP by comparing the unit atomic norm ball w.r.t. the atomic set of multi-view data points in a cluster and that w.r.t. the atomic set of remaining multi-view data points.
3. Based on the proposed ARP theory, we derive novel theoretical results for multi-view subspace clustering and classification, respectively.

1.4 Paper Organization

The remainder of the paper is arranged as below. Section 2 introduces some background knowledge and formulates the problem of multi-view subspace-preserving recovery. Section 3 presents the main results along with the comparison with other prior works. In Sections 4-5, we apply the proposed theory for multi-view subspace clustering and classification, respectively. Finally, Section 6 concludes.

2 Preliminaries and Problem Formulation

2.1 Notations

For better readability, scalars, vectors, matrices and sets are represented with italic letters (e.g., x), boldface lowercase letters (e.g., $\mathbf{x}$), boldface capital letters (e.g., $\mathbf{X}$) and calligraphic capital letters (e.g., $\mathcal{X}$), respectively. Denote by $[N]$ the set of integers from 1 to N. For any vector $\mathbf{x} \in \mathbb{R}^N$, supp($\mathbf{x}$) denotes its support set, i.e., $\text{supp}(\mathbf{x}) = \{i : x_i \neq 0\}$. Analogously, for any matrix $\mathbf{X} \in \mathbb{R}^{N \times V}$, rowsupp($\mathbf{X}$) denotes its row support set, i.e., the index set of nonzero rows of $\mathbf{X}$. Denote $\mathbf{X}(i,:)$ and $\mathbf{X}(:,j)$ as the i-th row and j-th column of $\mathbf{X}$, respectively. The $\ell_{1,2}$ norm of $\mathbf{X}$ is defined as $\|\mathbf{X}\|_{1,2} = \sum_{i=1}^{N} \|\mathbf{X}(i,:)\|_2$. To distinguish multi-view data from single-view data, a dot is added on the above of the variable (e.g., $\dot{\mathbf{A}}$). Table 1 summarizes the key notations and acronyms used in this paper.

Notation	Description
x, $\mathbf{x}$, $\mathbf{X}$, $\mathcal{X}$	scalar, vector, matrix, set
N	number of samples
$[N]$	set of integers from 1 to N
V	number of views
K	number of classes
d_v	data dimension of the v-th view
$\mathcal{S}_k^v \subset \mathbb{R}^{d_v}$	the k-th subspace in the v-th view
$\dot{\mathcal{S}}_k$	$\{\mathcal{S}_k^v, \forall v \in [V]\}$
$\mathbf{A}^v \in \mathbb{R}^{d_v \times N}$	data matrix of the v-th view
$\mathcal{A}^v$	set of columns of $\mathbf{A}^v$ and $-\mathbf{A}^v$
$\dot{\mathcal{A}}$	$\{\mathcal{A}^v, \forall v \in [V]\}$
$\mathcal{I}_k \subset [N]$	index set of samples in the k-th class
$\mathcal{I}_{-k} \subset [N]$	index set of samples out of the k-th class

Table 1: Key notations used in this paper.

2.2 Single-view Subspace-Preserving Recovery

Let $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \in \mathbb{R}^{d \times N}$ denote a matrix with nonzero columns collected from a union of K subspaces $\bigcup_{k=1}^{K} \mathcal{S}_k$, where each subspace corresponds to one class. Given any new nonzero point $\mathbf{y} \in \mathcal{S}_k$ for some k, subspace-preserving recovery aims to find a representation $\mathbf{c} \in \mathbb{R}^N$ of $\mathbf{y}$ such that (1) $\mathbf{y} = \mathbf{A}\mathbf{c}$, and (2) the nonzero entries of $\mathbf{c}$ only correspond to the columns of $\mathbf{A}$ lying in the same subspace as $\mathbf{y}$. Such a representation $\mathbf{c}$ is dubbed a subspace-preserving representation (SPR) [You and Vidal, 2015], which are formally defined as follows.

Definition 1. *(Subspace-Preserving Representation, SPR) [You and Vidal, 2015] Given $\mathbf{y} \in \mathcal{S}_k$, a solution $\mathbf{c}$ to $\mathbf{y} = \mathbf{A}\mathbf{c}$ is said to be a subspace-preserving representation (SPR) of $\mathbf{y}$ if and only if supp($\mathbf{c}$) $\subset \mathcal{I}_k$.*

Due to the promising performance and theoretical guarantee, sparse representation based subspace classification and clustering methods [Wright *et al.*, 2009; Elhamifar and Vidal, 2013] have attracted massive interest in recent years. Their success also motivate a plenty of theoretical works [You and Vidal, 2015; You *et al.*, 2016; Wang and Xu, 2016; Kaba *et al.*, 2021] to study conditions for sparse representation algorithms such as BP and OMP to give SPR solutions.

2.3 Multi-view Subspace-Preserving Recovery

Now we consider the setting of multi-view data. Assume that there are V views and K classes in total. For each view $v \in [V]$, let $\mathbf{A}^v = [\mathbf{a}_1^v, \cdots, \mathbf{a}_N^v] \in \mathbb{R}^{d_v \times N}$ denote a matrix with nonzero columns collected from a union of K subspaces $\bigcup_{k=1}^{K} \mathcal{S}_k^v$ where d_v is the v-th view data dimension. For simplicity, let $\dot{\mathbf{A}} = \{\mathbf{A}^v, \forall v \in [V]\}$. Let $\mathcal{I}_k$ and $\mathcal{I}_{-k}$ denote the index sets of samples in and out the k-th class. Let $\mathbf{A}_k^v$ and $\mathbf{A}_{-k}^v$ denote the sub-matrices of $\mathbf{A}^v$

containing and excluding samples from the subspace $\mathcal{S}_k^v$ for $k \in [K]$ and $v \in [V]$, respectively. For each class $k \in [K]$, let $\dot{\mathbf{A}}_k = \{\mathbf{A}_k^v, \forall v \in [V]\}$, $\dot{\mathbf{A}}_{-k} = \left\{\mathbf{A}_{-k}^v, \forall v \in [V]\right\}$, and $\dot{\mathcal{S}}_k = \{\mathcal{S}_k^v, \forall v \in [V]\}$. For simplicity, we say $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k$ if $\mathbf{y}^v \in \mathcal{S}_k^v$ holds for each view $v \in [V]$.

To leverage the correlation among multiple views, [Shekhar *et al.*, 2013] tackles the multi-view subspace classification by solving the joint sparse optimization program

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

where $\|\mathbf{C}\|_{1,2} = \sum_{i=1}^m \|\mathbf{C}(i,:)\|_2$ encourages the representation vector in different views to share the same sparsity pattern. To understand the correctness of this approach, we first propose the Multi-view SPR and consider the following theoretical questions.

Definition 2. *(Multi-view Subspace-Preserving Representation, MSPR) Given a multi-view sample $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k$, a representation matrix $\mathbf{C}$ satisfying $\mathbf{y}^v = \mathbf{A}^v \mathbf{C}(:, v)$, $\forall v \in [V]$ is said to be a Multi-view Subspace-Preserving Representation (MSPR) of $\dot{\mathbf{y}}$ if and only if rowsupp$(\mathbf{C}) \subset \mathcal{I}_k$.*

1. What are the theoretical conditions under which the solution to the model (1) is a MSPR?
2. Is there any geometric interpretable characterization of theoretical conditions for (1) to produce MSPR?
3. How to apply the theoretical results on MSPR for multi-view subspace clustering and classification?

3 Main Results

In this section, we aim to provide answers to the questions concerning MSPR in Section 2. Before showing the main results, we first introduce the definition of atomic norm and extend it for multi-view data.

3.1 Multi-view Atomic Norm

Let $\mathcal{A}$ be an origin-symmetric (i.e., $\mathbf{a} \in \mathcal{A}$ *if and only if* (iff) $-\mathbf{a} \in \mathcal{A}$) set. The definition of atomic norm is introduced as follows.

Definition 3. *[Chandrasekaran* et al.*, 2012] Given the atomic set $\mathcal{A}$, the atomic norm of $\mathbf{x}$ with respect to $\mathcal{A}$ is defined as*

$$\|\mathbf{x}\|_{\mathcal{A}} := \inf \{t > 0 : \mathbf{x} \in t \cdot conv(\mathcal{A})\},$$

where $conv(\mathcal{A})$ is the convex hull of the set $\mathcal{A}$.

If the atomic set $\mathcal{A}$ is centrally symmetric about the origin, the gauge function $\|\cdot\|_{\mathcal{A}}$ is a norm [Chandrasekaran *et al.*, 2012]. In this case, we have

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf_{c_{\mathbf{a}} \geq 0, \forall \mathbf{a} \in \mathcal{A}} \left\{ \sum_{\mathbf{a} \in \mathcal{A}} c_{\mathbf{a}} : \mathbf{x} = \sum_{\mathbf{a} \in \mathcal{A}} c_{\mathbf{a}} \mathbf{a} \right\}. \quad (2)$$

Many popular norms such as the ℓ_1 norm and matrix nuclear norm are special cases of atomic norm with appropriate choice of the atomic set [Chandrasekaran *et al.*, 2012]. The dual norm of $\|\cdot\|_{\mathcal{A}}$ is defined by

$$\|\mathbf{x}\|_{\mathcal{A}}^* := \sup \{\langle \mathbf{x}, \mathbf{a} \rangle, \mathbf{a} \in \mathcal{A}\},$$

where $\langle \mathbf{x}, \mathbf{a} \rangle = \text{tr}(\mathbf{x}^T \mathbf{a})$ and $\text{tr}(\cdot)$ represents the trace of a matrix. Now we generalize the atomic norm to the multi-view setting.

Definition 4. *Given the multi-view atomic set $\dot{\mathcal{A}} = \{\mathcal{A}^v\}_{v=1}^V$, the Multi-view Atomic Norm (MAN) of the multi-view data $\dot{\mathbf{x}} = \{\mathbf{x}^v\}_{v=1}^V$ with respect to $\dot{\mathcal{A}}$ is defined as*

$$\|\dot{\mathbf{x}}\|_{\dot{\mathcal{A}}} = \inf \left\{ \sum_{\dot{\mathbf{a}} \in \dot{\mathcal{A}}} \|\mathbf{c}_{\dot{\mathbf{a}}}\|_2 : \mathbf{x}^v = \sum_{\mathbf{a}^v \in \mathcal{A}^v} \mathbf{c}_{\dot{\mathbf{a}}}(v) \mathbf{a}^v, v \in [V] \right\}.$$

Remark 1. *Note that when $V = 1$, the multi-view atomic norm reduces to the atomic norm in Definition 3.*

3.2 Atomic Recovery Property

Before showing the theoretical results, we first introduce some useful lemmas.

Lemma 1. *If $\dot{\mathcal{A}}_1 \subset \dot{\mathcal{A}}_2$, $\|\dot{\mathbf{x}}\|_{\dot{\mathcal{A}}_1} \geq \|\dot{\mathbf{x}}\|_{\dot{\mathcal{A}}_2}$.*

Proof. For any $t > 0$ such that $\mathbf{y} \in t \cdot \text{conv}(\dot{\mathcal{A}}_1)$, we have $\mathbf{y} \in t \cdot \text{conv}(\dot{\mathcal{A}}_2)$. So

$$\left\{ t : \mathbf{y} \in t \cdot \text{conv}(\dot{\mathcal{A}}_1) \right\} \subset \left\{ t : \mathbf{y} \in t \cdot \text{conv}(\dot{\mathcal{A}}_2) \right\}.$$

Recalling the definition of MAN gives $\|\mathbf{y}\|_{\dot{\mathcal{A}}_1} \geq \|\mathbf{y}\|_{\dot{\mathcal{A}}_2}$. □

Recall that $\mathbf{A}_k^v$ is the submatrix of $\mathbf{A}^v$ including samples from the k-th subspace $\mathcal{S}_k^v$ of the v-th view and $\mathbf{A}_{-k}^v$ is the submatrix of $\mathbf{A}^v$ by excluding the samples in $\mathcal{S}_k^v$. Let $\mathcal{A}_k^v$ and $\mathcal{A}_{-k}^v$ denote the set of columns of $\pm \mathbf{A}_k^v$ and $\pm \mathbf{A}_{-k}^v$, respectively. We can define the multi-view atomic sets $\dot{\mathcal{A}}_k = \{\mathcal{A}_k^v\}_{v=1}^V$ and $\dot{\mathcal{A}}_{-k} = \{\mathcal{A}_{-k}^v\}_{v=1}^V$. Then we have the following results to characterize the relationship among the three multi-view atomic norms, i.e., $\|\cdot\|_{\dot{\mathcal{A}}}$, $\|\cdot\|_{\dot{\mathcal{A}}_k}$ and $\|\cdot\|_{\dot{\mathcal{A}}_{-k}}$.

Lemma 2. *For any optimal solution $\mathbf{C}^*$ to the model (1), let $\dot{\mathbf{y}}_k$ and $\dot{\mathbf{y}}_{-k}$ are defined such that $\mathbf{y}_k^v = \mathbf{A}_k^v \mathbf{C}^*(\mathcal{I}_k, :)$ and $\mathbf{y}_{-k}^v = \mathbf{A}_{-k}^v \mathbf{C}^*(\mathcal{I}_{-k}, :)$ for $v = 1, \cdots, V$. Then we have*

$$\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}} = \|\dot{\mathbf{y}}_k\|_{\dot{\mathcal{A}}_k} + \|\dot{\mathbf{y}}_{-k}\|_{\dot{\mathcal{A}}_{-k}}. \quad (3)$$

Proof. On the one hand, since $\dot{\mathcal{A}}_k$, $\dot{\mathcal{A}}_{-k} \subset \dot{\mathcal{A}}$ and by Lemma 1, there holds

$$\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}} \leq \|\dot{\mathbf{y}}_k\|_{\dot{\mathcal{A}}} + \|\dot{\mathbf{y}}_{-k}\|_{\dot{\mathcal{A}}} \leq \|\dot{\mathbf{y}}_k\|_{\dot{\mathcal{A}}_k} + \|\dot{\mathbf{y}}_{-k}\|_{\dot{\mathcal{A}}_{-k}}. \quad (4)$$

On the other hand, according to the definition of the Multi-view Atomic Norm (MAN) we have

$$\|\mathbf{y}_k\|_{\dot{\mathcal{A}}_k} \leq \|\mathbf{C}^*(\mathcal{I}_k, :)\|_{1,2}, \quad (5)$$

$$\|\dot{\mathbf{y}}_{-k}\|_{\dot{\mathcal{A}}_{-k}} \leq \|\mathbf{C}^*(\mathcal{I}_{-k}, :)\|_{1,2}. \quad (6)$$

It follows that

$$\begin{aligned} &\|\dot{\mathbf{y}}_k\|_{\dot{\mathcal{A}}_k} + \|\dot{\mathbf{y}}_{-k}\|_{\dot{\mathcal{A}}_{-k}} \\ &\leq \|\mathbf{C}^*(\mathcal{I}_k, :)\|_{1,2} + \|\mathbf{C}^*(\mathcal{I}_{-k}, :)\|_{1,2} \\ &= \|\mathbf{C}^*\|_{1,2} = \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}}. \end{aligned} \quad (7)$$

Combining (4) and (7) completes the proof. □

Now we give the definition of ARP for the multi-view data $\dot{\mathbf{A}}$, which is beneficial for producing MSPR.

Definition 5. *(Atomic Recovery Property, ARP) The multi-view data $\dot{\mathbf{A}}$ is said to satisfy the Atomic Recovery Property (ARP) if for all $k \in [K]$ and $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$, we have*

$$\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} < \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}}. \tag{8}$$

Remark 2. *The ARP means that it is more* concise *in terms of the atomic norm to represent any nonzero sample $\dot{\mathbf{y}}$ in $\dot{\mathcal{S}}_k$ by using $\dot{\mathcal{A}}_k$ than $\dot{\mathcal{A}}_{-k}$.*

With the definitions above, we have the following result.

Theorem 1. *The solution to (1) is a MSPR for all $k \in [K]$ and $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$ if and only if $\dot{\mathbf{A}}$ satisfies the ARP.*

Proof. First, we prove that any solution $\mathbf{C}^* = \mathbf{C}^*\left(\dot{\mathbf{y}}, \dot{\mathbf{A}}\right)$ to (1) is a MSPR for any $k \in [K]$ and $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$ if $\dot{\mathbf{A}}$ satisfies the ARP. That is, we aim to prove $\mathbf{C}^*(\mathcal{I}_{-k}, :) = \mathbf{0}$. Denote

$$\mathbf{r}^v := \mathbf{y}^v - \mathbf{A}_k^v \mathbf{C}^*(\mathcal{I}_k, v) = \mathbf{A}_{-k}^v \mathbf{C}^*(\mathcal{I}_{-k}, v). \tag{9}$$

The second equality above uses the fact that $\mathbf{y}^v = \mathbf{A}^v \mathbf{C}^*(:, v) = \mathbf{A}_k^v \mathbf{C}^*(\mathcal{I}_k, v) + \mathbf{A}_{-k}^v \mathbf{C}^*(\mathcal{I}_{-k}, v)$. The proof consists of the following two steps

1) **Step 1:** If $\mathbf{r}^v = \mathbf{0}, \forall v \in [V]$, $\mathbf{C}^*(\mathcal{I}_{-k}, :) = \mathbf{0}$.
2) **Step 2:** Prove $\mathbf{r}^v = \mathbf{0}, \forall v \in [V]$ by contradiction.

Proof of Step 1: For any $v \in [V]$, since $\mathbf{y}^v \in \mathcal{S}_k^v$ and $\mathbf{A}_k^v \mathbf{C}^*(\mathcal{I}_k, v) \in \mathcal{S}_k^v$, we have $\mathbf{r}^v \in \mathcal{S}_k^v$ according to Eq. (9). If $\mathbf{r}^v = \mathbf{0}$, $\mathbf{y}^v = \mathbf{A}_k^v \mathbf{C}^*(\mathcal{I}_k, v) = \mathbf{A}\delta_k(\mathbf{C}^*(:, v))$. Let $\Delta_k(\mathbf{C}^*)$ be the matrix by arranging $\delta_k(\mathbf{C}^*(:, v))$ as its v-th column for any $v \in [V]$. Then $\Delta_k(\mathbf{C}^*)$ is also feasible to the JSR model (1). If $\Delta_{-k}(\mathbf{C}^*) \neq \mathbf{0}$, we have

$$\|\Delta_k(\mathbf{C}^*)\|_1 < \|\mathbf{C}^*\|_1 .$$

This contradicts with the optimality of $\mathbf{C}^*$ to (1). Thus $\Delta_{-k}(\mathbf{C}^*) = \mathbf{0}$. So if we can prove $\mathbf{r}^v = \mathbf{0}$ for any $v \in [V]$, the desired conclusion holds. Now we prove it by contradiction.

Proof of Step 2: If $\dot{\mathbf{r}} \neq \dot{\mathbf{0}}$, since $\dot{\mathbf{A}}$ satisfies the ARP and $\dot{\mathbf{r}} \in \dot{\mathcal{S}}_k$, there holds

$$\|\dot{\mathbf{r}}\|_{\dot{\mathcal{A}}_k} < \|\dot{\mathbf{r}}\|_{\dot{\mathcal{A}}_{-k}} . \tag{10}$$

For ease of simplicity, let $\mathbf{y}_k^v = \mathbf{A}_k^v \mathbf{C}^*(\mathcal{I}_k, v), \forall v \in [V]$. Accordingly, $\mathbf{y}^v = \mathbf{y}_k^v + \mathbf{r}^v, \forall v \in [V]$. It follows that

$$\begin{aligned}\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}} &= \|\dot{\mathbf{y}}_k + \dot{\mathbf{r}}\|_{\dot{\mathcal{A}}} \\ &\leq \|\dot{\mathbf{y}}_k + \dot{\mathbf{r}}\|_{\dot{\mathcal{A}}_k} \text{ (By Lemma 1)} \\ &\leq \|\dot{\mathbf{y}}_k\|_{\dot{\mathcal{A}}_k} + \|\dot{\mathbf{r}}\|_{\dot{\mathcal{A}}_k} \text{ (By triangle inequality)} \\ &< \|\dot{\mathbf{y}}_k\|_{\dot{\mathcal{A}}_k} + \|\dot{\mathbf{r}}\|_{\dot{\mathcal{A}}_{-k}} \text{ (By the ARP property)} \\ &= \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}}. \text{ (By Lemma 2)}\end{aligned}$$

This is a contradiction. So $\dot{\mathbf{r}} = \mathbf{0}$ and $\mathbf{C}^*$ is subspace-preserving.

Conversely, if any solution $\mathbf{C}^*$ to (1) is a MSPR for any $k \in [K]$ and $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\mathbf{0}\}$, we will prove $\dot{\mathbf{A}}$ satisfies the ARP by contradiction. Assume that there exists some $\dot{\mathbf{y}}$ violating the ARP, i.e., $\|\dot{\mathbf{y}}\|_{\mathcal{A}_k} \geq \|\dot{\mathbf{y}}\|_{\mathcal{A}_{-k}}$. Consider the following two optimization problems

$$\mathbf{C}_k^* = \underset{\mathbf{C} \in \mathbb{R}^{N_k \times V}}{\arg\min} \|\mathbf{C}\|_{1,2} \text{ s.t. } \mathbf{y}^v = \mathbf{A}_k^v \mathbf{C}(:, v), \forall v \in [V], \tag{11}$$

$$\mathbf{C}_{-k}^* = \underset{\mathbf{C} \in \mathbb{R}^{N_{-k} \times V}}{\arg\min} \|\mathbf{C}\|_{1,2} \text{ s.t. } \mathbf{y}^v = \mathbf{A}_{-k}^v \mathbf{C}(:, v), \forall v \in [V]. \tag{12}$$

By the hypothesis we know that $\mathbf{C}^*$ is a MSPR and $\text{rowsupp}(\mathbf{C}^*) \subset \mathcal{I}_k$. So $\mathbf{C}^*(\mathcal{I}_k, :)$ is also an optimal solution to the problem (11). It follows that

$$\begin{aligned}\|\mathbf{C}^*\|_{1,2} &= \|\mathbf{C}^*(\mathcal{I}_k, :)\|_{1,2} = \|\mathbf{C}_k^*\|_{1,2} \\ &= \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} \geq \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}} = \left\|\mathbf{C}_{-k}^*\right\|_{1,2}.\end{aligned}$$

Define a matrix $\tilde{\mathbf{C}} \in \mathbb{R}^{N \times V}$ such that $\tilde{\mathbf{C}}(\mathcal{I}_k, :) = \mathbf{0}$ and $\tilde{\mathbf{C}}(\mathcal{I}_{-k}, :) = \mathbf{C}_{-k}^*$. Then we have $\|\tilde{\mathbf{C}}\|_{1,2} \leq \|\mathbf{C}_{-k}^*\|_{1,2} = \|\mathbf{C}^*\|_{1,2}$ and $\mathbf{y}^v = \mathbf{A}^v \tilde{\mathbf{C}}(:, v), \forall v \in [V]$. Thus $\tilde{\mathbf{C}}$ is feasible to (1) and also an optimal solution to (1). However, according to the definitions of $\mathbf{C}_{-k}^*$ in (12) and $\tilde{\mathbf{C}}$, we know that $\tilde{\mathbf{C}}$ is not a MSPR. This contradicts with the hypothesis that any solution to (1) is a MSPR. This completes the proof. □

Remark 3. *Theorem 1 shows that ARP is a necessary and sufficient condition for the model (1) to generate MSPR.*

3.3 Geometric Characterization of ARP

The unit ball with respect to the atomic set $\dot{\mathcal{A}}$ is defined as $\mathbb{B}_{\dot{\mathcal{A}}} := \{\dot{\mathbf{y}} : \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}} \leq 1\}$.

Theorem 2. *The multi-view data $\dot{\mathbf{A}}$ satisfies the ARP if and only if for all $k \in [K]$, there holds*

$$\dot{\mathcal{S}}_k \cap \mathbb{B}_{\dot{\mathcal{A}}_{-k}} \subseteq \text{rinte}\left(\mathbb{B}_{\dot{\mathcal{A}}_k}\right), \tag{13}$$

where rinte $\left(\mathbb{B}_{\dot{\mathcal{A}}_k}\right)$ *denotes the relative interior of* $\mathbb{B}_{\dot{\mathcal{A}}_k}$.

Proof. Firstly, we prove that $\dot{\mathbf{A}}$ satisfies the ARP if Eq. (13) holds for all $k \in [K]$. For all $k \in [K]$ and $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$, define $\dot{\mathbf{z}}$ such that $\mathbf{z}^v = \frac{\mathbf{y}^v}{\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}}}$ for $v \in [V]$. It follows that $\dot{\mathbf{z}} \in \dot{\mathcal{S}}$ and $\|\dot{\mathbf{z}}\|_{\dot{\mathcal{A}}_{-k}} = 1$. Thus, $\dot{\mathbf{z}} \in \dot{\mathcal{S}}_k \cap \mathbb{B}_{\dot{\mathcal{A}}_{-k}} \subseteq$ rinte $\left(\mathbb{B}_{\dot{\mathcal{A}}_k}\right)$. Accordingly, $\|\dot{\mathbf{z}}\|_{\dot{\mathcal{A}}_k} < 1$. Recall the definition of $\dot{\mathbf{z}}$, this is equivalent to $\left\|\frac{\dot{\mathbf{y}}}{\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}}}\right\|_{\dot{\mathcal{A}}_k} < 1$. This gives $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} < \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}}$.

Conversely, we prove that if $\dot{\mathbf{A}}$ satisfies the ARP, Eq. (13) holds for all $k \in [K]$. For any $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \cap \mathbb{B}_{\dot{\mathcal{A}}_{-k}}$, $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} < \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}} \leq 1$. Therefore, we have $\dot{\mathbf{y}} \in \mathbb{B}_{\dot{\mathcal{A}}_k}$. □

For single-view data, we have the following result.

Corollary 1. *The single-view data $\mathbf{A}$ satisfies the ARP if and only if for all $k \in [K]$, there holds*

$$\mathcal{S}_k \cap \mathbb{B}_{\mathcal{A}_{-k}} \subseteq \textit{rinte}\left(\mathbb{B}_{\mathcal{A}_k}\right). \tag{14}$$

Lemma 3. *Let $\mathcal{A}$ be the set of columns of $\mathbf{A}$ and $-\mathbf{A}$. We have $\mathbb{B}_{\mathcal{A}} = \mathcal{K}(\mathbf{A})$.*

Proof. We first prove that $\mathbb{B}_{\mathcal{A}} \subset \mathcal{K}(\mathbf{A})$. $\forall \mathbf{y} \in \mathbb{B}_{\mathcal{A}}$, we have $\|\mathbf{y}\|_{\mathcal{A}} = 1$. According to Eq. (2), there exists $\{c_{\mathbf{a}}, \mathbf{a} \in \mathcal{A}\}$ such that $\mathbf{y} = \sum_{\mathbf{a}\in\mathcal{A}} c_{\mathbf{a}}\mathbf{a}, c_{\mathbf{a}} \geq 0$ and $\sum_{\mathbf{a}\in\mathcal{A}} c_{\mathbf{a}} = 1$. Thus, $\mathbf{y}$ lies in the convex hull of $\pm\mathbf{A}$ and we have $\mathbf{y} \in \mathcal{K}(\mathbf{A})$.

Then we prove $\mathcal{K}(\mathbf{A}) \subset \mathbb{B}_{\mathcal{A}}$. $\forall \mathbf{y} \in \mathcal{K}(\mathbf{A})$, there exists $\{\lambda_{\mathbf{a}}\}_{\mathbf{a}\in\mathcal{A}}$ ($\lambda_{\mathbf{a}} \geq 0$) such that $\mathbf{y} = \sum_{\mathbf{a}\in\mathcal{A}} \lambda_{\mathbf{a}}\mathbf{a}$ and $\sum_{\mathbf{a}\in\mathcal{A}} \lambda_{\mathbf{a}} = 1$. According to the definition of atomic norm (2), there holds $\|\mathbf{y}\|_{\mathcal{A}} \leq 1$ and $\mathbf{y} \in \mathbb{B}_{\mathcal{A}}$. □

Corollary 2. *The single-view data* $\mathbf{A}$ *satisfies the ARP if and only if for all* $k \in [K]$*, there holds*

$$\mathcal{S}_k \cap \mathcal{K}(\mathbf{A}_{-k}) \subseteq \textit{rinte}\left(\mathcal{K}(\mathbf{A}_k)\right). \tag{15}$$

Here $\mathcal{K}(\mathbf{A}) = \mathrm{conv}(\pm\mathbf{A})$ denotes the symmetrized convex hull of the columns of $\mathbf{A}$ and $\mathrm{rinte}\left(\mathcal{K}(\mathbf{A}_k)\right)$ is the relative interior of $\mathcal{K}(\mathbf{A}_k)$. Therefore, Theorem 2 for single-view data in the previous work [Kaba *et al.*, 2021] can be regarded as a special case of our result in Theorem 2 when $V = 1$.

Next, we derive another geometric characterization for ARP based on the multi-view incoherence and multi-view circumradius, which are defined as follows.

Definition 6. ***(Multi-view Incoherence)*** *The multi-view incoherence between two multi-view datasets* $\dot{\mathbf{P}} = \{\mathbf{P}^v\}_{v=1}^V$ *and* $\dot{\mathbf{Q}} = \{\mathbf{Q}^v\}_{v=1}^V$ *is defined as*

$$\mu_M\left(\dot{\mathbf{P}}, \dot{\mathbf{Q}}\right) := \max_{\dot{\mathbf{p}}\in\dot{\mathbf{P}}, \dot{\mathbf{q}}\in\dot{\mathbf{Q}}} \left\| \left[\frac{\langle \mathbf{p}^1, \mathbf{q}^1\rangle}{\|\mathbf{p}^1\|_2\|\mathbf{q}^1\|_2}, \cdots, \frac{\langle \mathbf{p}^V, \mathbf{q}^V\rangle}{\|\mathbf{p}^V\|_2\|\mathbf{q}^V\|_2} \right] \right\|_2. \tag{16}$$

where $\dot{\mathbf{p}} = \{\mathbf{p}^v\}_{v=1}^V$ *and* $\dot{\mathbf{q}} = \{\mathbf{q}^v\}_{v=1}^V$.

Remark 4. *By substituting* $\dot{\mathbf{A}}_{-k}$ *and* $\dot{\mathcal{S}}_k$ *into Eq. (16), we have*

$$\mu_M\left(\dot{\mathbf{A}}_{-k}, \dot{\mathcal{S}}_k\right) := \max_{\mathbf{u}^v\in\mathcal{S}_k^v\cap\mathbb{S}^{d_v-1}} \left\| \left[\left(\mathbf{A}_{-k}^1\right)^T \mathbf{u}^1, \cdots, \left(\dot{\mathbf{A}}_{-k}^V\right)^T \mathbf{u}^V \right] \right\|_{\infty,2}, \tag{17}$$

where $\mathbb{S}^{d_v-1} = \left\{\mathbf{z} \in \mathbb{R}^{d_v} : \|\mathbf{z}\|_2 = 1\right\}$ *is the unit sphere in* $\mathbb{R}^{d_v}$. *The multi-view incoherence* $\mu_M\left(\dot{\mathbf{A}}_{-k}, \dot{\mathcal{S}}_k\right)$ *measures how close the multi-view samples in other clusters from the underlying subspaces corresponding to the* k*-th cluster.*

Definition 7. ***(Multi-view Circumradius)*** *The multi-view circumradius of* $\dot{\mathbf{A}}_k = \{\mathbf{A}_k^v\}_{v=1}^V$ *is defined as*

$$\begin{aligned} R_M\left(\dot{\mathbf{A}}_k\right) &= \max_{\dot{\mathbf{u}}\in\dot{\mathcal{S}}_k} \max_{v\in[V]} \|\mathbf{u}^v\|_2, \\ \text{s.t. } &\left\| \left[\left(\mathbf{A}_k^1\right)^T \mathbf{u}^1, \cdots, \left(\mathbf{A}_k^V\right)^T \mathbf{u}^V \right] \right\|_{\infty,2} \leq 1, \end{aligned} \tag{18}$$

where $\dot{\mathbf{u}} = \{\mathbf{u}^v\}_{v=1}^V$.

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

Theorem 3. *The multi-view data* $\dot{\mathbf{A}}$ *satisfies the ARP if for all* $k \in [K]$*, there holds*

$$\mu_M\left(\dot{\mathbf{A}}_{-k}, \dot{\mathcal{S}}_k\right) R_M\left(\dot{\mathbf{A}}_k\right) < 1,\ k = 1, \cdots, K. \tag{19}$$

Proof. For any $k \in [K]$ and $\dot{\mathbf{y}} \in \dot{\mathcal{S}}_k \backslash \{\dot{\mathbf{0}}\}$, we prove that $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} < \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}}$. The proof has the following two steps

1) **Step 1:** $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} \leq \left(\sum_{v=1}^V \|\mathbf{y}^v\|_2\right) R_M\left(\dot{\mathbf{A}}_k\right)$.
2) **Step 2:** $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}} \geq \left(\sum_{v=1}^V \|\mathbf{y}^v\|_2\right) / \mu_M\left(\dot{\mathbf{A}}_{-k}, \dot{\mathcal{S}}_k\right)$.

We first consider the following two row-sparse optimization problems

$$\mathbf{C}_k^* = \underset{\mathbf{C}\in\mathbb{R}^{N_k\times V}}{\arg\min} \|\mathbf{C}\|_{1,2} \text{ s.t. } \mathbf{y}^v = \mathbf{A}_k^v\mathbf{C}(:,v),\ \forall v \in [V], \tag{20}$$

$$\mathbf{C}_{-k}^* = \underset{\mathbf{C}\in\mathbb{R}^{N_{-k}\times V}}{\arg\min} \|\mathbf{C}\|_{1,2} \text{ s.t. } \mathbf{y}^v = \mathbf{A}_{-k}^v\mathbf{C}(:,v),\ \forall v \in [V], \tag{21}$$

Proof of Step 1: Since $\mathbf{y}^v \in \mathcal{S}_k^v = \mathrm{span}(\mathbf{A}_k)$ for each $v \in [V]$, the problem (20) is feasible and the optimal solution $\mathbf{C}_k^*$ exists. The dual problem of (20) can be formulated as

$$\begin{aligned} \dot{\mathbf{u}}_k &= \underset{\dot{\mathbf{u}}=\{\mathbf{u}^v\}_{v=1}^V}{\arg\max} \sum_{v=1}^V \langle \mathbf{u}^v, \mathbf{y}^v\rangle \\ \text{s.t. } &\left\| \left[\left(\mathbf{A}_k^1\right)^T \mathbf{u}^1, \cdots, \left(\mathbf{A}_k^V\right)^T \mathbf{u}^V \right] \right\|_{\infty,2} \leq 1. \end{aligned} \tag{22}$$

For each view $v \in [V]$, decompose $\mathbf{u}_k^v$ into two orthogonal components $\mathbf{u}_k^v = \mathbf{u}_k^{v,\|} + \mathbf{u}_k^{v,\perp}$, where $\mathbf{u}^{v,\|} \in \mathcal{S}_k^v$ and $\mathbf{u}_k^{v,\perp} \in \mathcal{S}_k^{v,\perp}$. Here $\mathcal{S}_k^{v,\perp}$ denotes the orthogonal complement of $\mathcal{S}_k^v$ in $\mathbb{R}^{d_v}$. Then we have $\left(\mathbf{A}_k^v\right)^T \mathbf{u}_k^v = \left(\mathbf{A}_k^v\right)^T \mathbf{u}_k^{v,\|}$ for any $v \in [V]$. It follows that

$$\begin{aligned} &\left\| \left[\left(\mathbf{A}_k^1\right)^T \mathbf{u}_k^{1,\|}, \cdots, \left(\mathbf{A}_k^V\right)^T \mathbf{u}_k^{V,\|} \right] \right\|_{\infty,2} \\ =&\left\| \left[\left(\mathbf{A}_k^1\right)^T \mathbf{u}_k^1, \cdots, \left(\mathbf{A}_k^V\right)^T \mathbf{u}_k^V \right] \right\|_{\infty,2} \leq 1. \end{aligned} \tag{23}$$

The second equality above uses the fact that $\mathbf{u}_k^{\perp}$ is orthogonal to $\mathcal{S}_k^v = \mathrm{span}(\mathbf{A}_k^v)$. Let $\dot{\mathbf{u}}_k^{\|} = \left\{\mathbf{u}_k^{v,\|}\right\}_{v=1}^V$. Since $\mathbf{u}_k^{v,\|} \in \mathcal{S}_k^v$, $\forall v \in [V]$, there holds $\dot{\mathbf{u}}_k^{\|} \in \dot{\mathcal{S}}_k$. According to the definition of $R_M\left(\dot{\mathbf{A}}_k\right)$ in Eq. (18), we have

$$\max_{v\in[V]} \left\|\mathbf{u}_k^{v,\|}\right\|_2 \leq R_M\left(\dot{\mathbf{A}}_k\right). \tag{24}$$

By the strong duality of (20) and (22) and noting $\mathbf{y}^v \in \mathcal{S}_k^v$, there holds

$$\begin{aligned} \|\mathbf{C}_k^*\|_{1,2} &= \sum_{v=1}^V \langle \mathbf{u}_k^v, \mathbf{y}^v\rangle = \sum_{v=1}^V \left\langle \mathbf{u}_k^{v,\|}, \mathbf{y}^v\right\rangle \\ &\leq \left(\sum_{v=1}^V \|\mathbf{y}^v\|_2\right) \max_{v\in[V]} \left\|\mathbf{u}_k^{v,\|}\right\|_2. \end{aligned} \tag{25}$$

Combining Eq. (24) and Eq. (25) gives $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} = \|\mathbf{C}_k^*\|_{1,2} \leq \left(\sum_{v=1}^V \|\mathbf{y}^v\|_2\right) R_M\left(\dot{\mathbf{A}}_k\right)$.

Proof of Step 2: If the problem (21) is not feasible, $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}} = \left\|\mathbf{C}^*_{-k}\right\|_{1,2} = +\infty$ and the desired conclusion holds. Otherwise, since $\mathbf{C}^*_{-k}$ is feasible to (21), we can write

$$\mathbf{y}^v = \mathbf{A}^v_{-k}\mathbf{C}^*_{-k}(:,v),\ \forall v \in [V].$$

Multiplying both sides of the above equation on the left by $(\mathbf{y}^v)^T/\|\mathbf{y}^v\|_2$, we have

$$\|\mathbf{y}^v\|_2 = (\mathbf{y}^v)^T/\|\mathbf{y}^v\|_2\mathbf{A}^v_{-k}\mathbf{C}^*_{-k}(:,v),\ \forall v \in [V].$$

It follows that

$$\begin{aligned}
&\sum_{v=1}^{V}\|\mathbf{y}^v\|_2 = \sum_{v=1}^{V}(\mathbf{y}^v)^T/\|\mathbf{y}^v\|_2\mathbf{A}^v_{-k}\mathbf{C}^*_{-k}(:,v)\\
=&\sum_{v=1}^{V}\left\langle \left(\mathbf{A}^v_{-k}\right)^T\frac{\mathbf{y}^v}{\|\mathbf{y}^v\|_2}, \mathbf{C}^*_{-k}(:,v)\right\rangle\\
=&\left\langle\left[\left(\mathbf{A}^1_{-k}\right)^T\frac{\mathbf{y}^1}{\|\mathbf{y}^1\|_2},\cdots,\left(\mathbf{A}^V_{-k}\right)^T\frac{\mathbf{y}^V}{\|\mathbf{y}^V\|_2}\right], \mathbf{C}^*_{-k}\right\rangle\\
\leq&\left\|\left[\left(\mathbf{A}^1_{-k}\right)^T\frac{\mathbf{y}^1}{\|\mathbf{y}^1\|_2},\cdots,\left(\mathbf{A}^V_{-k}\right)^T\frac{\mathbf{y}^V}{\|\mathbf{y}^V\|_2}\right]\right\|_{\infty,2}\left\|\mathbf{C}^*_{-k}\right\|_{1,2}\\
\leq&\mu_M\left(\dot{\mathbf{A}}_{-k},\dot{\mathcal{S}}_k\right)\left\|\mathbf{C}^*_{-k}\right\|_{1,2}.
\end{aligned} \tag{26}$$

Since $\mathbf{y}^v \neq \mathbf{0}$ and $\|\mathbf{y}^v\|_2 > 0$, the equation (26) above indicates that $\mu_M\left(\dot{\mathbf{A}}_{-k},\dot{\mathcal{S}}_k\right) > 0$ and

$$\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}} = \left\|\mathbf{C}^*_{-k}\right\|_{1,2} \geq \sum_{v=1}^{V}\|\mathbf{y}^v\|_2/\mu_M\left(\dot{\mathbf{A}}_{-k},\dot{\mathcal{S}}_k\right). \tag{27}$$

This finishes the proof of Step 2.

Combining Steps 1 and 2 and Eq. (19), we have $\|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_k} < \|\dot{\mathbf{y}}\|_{\dot{\mathcal{A}}_{-k}}$. This completes the proof of the theorem. □

3.4 Comparison with Prior Works

To further clarify the novelty and contributions of the paper, we compare our results with prior related works. Table 2 compares the important prior works and this work on SPR. Note that most prior works are concerned with the *single-view* data instead of *multi-view* data considered in this paper.

Comparison to [Soltanolkotabi and Candés, 2012]. In [Soltanolkotabi and Candés, 2012], the authors shows that the SR model Basis Pursuit (BP) can give SPR for a specific point $\mathbf{y}$ in $\mathcal{S}_k$ rather than all points in $\mathcal{S}_k$ if

$$\mu\left(\mathbf{A}_{-k},\mathbf{v}\left(\mathbf{y},\mathbf{A}_k\right)\right) < r\left(\mathcal{K}\left(\pm\mathbf{A}_k\right)\right), \tag{28}$$

where the dual direction $\mathbf{v}(\mathbf{y},\mathbf{A}_k)$ of $\mathbf{y}$ is defined as

$$\mathbf{v}(\mathbf{y},\mathbf{A}_k) = \mathbf{U}_k\frac{\boldsymbol{\lambda}(\mathbf{y},\mathbf{A}_k)}{\|\boldsymbol{\lambda}(\mathbf{y},\mathbf{A}_k)\|_2}.$$

Here $\boldsymbol{\lambda}(\mathbf{y},\mathbf{A}_k)$ is called the dual point and defined as an optimal solution with minimum Euclidean norm to the problem

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^m}\langle\mathbf{y},\boldsymbol{\lambda}\rangle,\ \text{ s.t. }\ \left\|\mathbf{A}_k^T\boldsymbol{\lambda}\right\|_\infty \leq 1.$$

We can apply it to all samples in $\mathcal{S}_k$ and get the condition

$$\mu\left(\mathbf{A}_{-k},\mathcal{V}\right) < r\left(\mathcal{K}\left(\pm\mathbf{A}_k\right)\right), \tag{29}$$

where $\mathcal{V} := \{\mathbf{v}\left(\mathbf{y},\mathbf{A}_k\right),\forall\mathbf{y}\in\mathcal{S}_k\}$. Although the condition above provides clear geometric insights for the success of SSC, it is only a sufficient condition for SPR. In addition, it only involves the single-view data.

Comparison to [You and Vidal, 2015]. In [You and Vidal, 2015], the authors show that the sparse representation models BP [Chen *et al.*, 1998] and OMP [Pati *et al.*, 1993] can yield subspace-preserving representation (SPR) under the following principal recovery condition (PRC)

$$r\left(\mathcal{K}\left(\pm\mathbf{A}_k\right)\right) > \mu\left(\mathbf{A}_{-k},\mathcal{R}(\mathbf{A}_k)\right) \tag{30}$$

where $r\left(\mathcal{K}\left(\pm\mathbf{A}_k\right)\right)$ is defined as the radius of the largest ball inscribed in $\mathcal{K}\left(\pm\mathbf{A}_k\right)$, which denotes the convex hull of the symmetrized columns of $\mathbf{A}_k$. Here $\mathcal{R}(\mathbf{A}_k)$ denotes the range of $\mathbf{A}_k$, i.e., the linear span of columns of $\mathbf{A}_k$. The incoherence $\mu\left(\mathbf{A}_{-k},\mathcal{R}(\mathbf{A}_k)\right)$ is defined as

$$\mu\left(\mathbf{A}_{-k},\mathcal{R}(\mathbf{A}_k)\right) = \max_{\mathbf{z}\in\mathcal{R}(\mathbf{A}_k)\setminus\{\mathbf{0}\}}\left\|\mathbf{A}_{-k}^T\mathbf{z}\right\|_\infty/\|\mathbf{z}\|_2.$$

While the authors developed the dual recovery condition (DRC) to relax PRC, both DRC and PRC are only sufficient conditions and are confined to single-view data.

Comparison to [Kaba *et al.*, 2021]. In [Kaba *et al.*, 2021], the authors also considers the SPR problem and derives a necessary and sufficient condition termed subspace nullspace property (SNSP) for SPR. Specifically, the SNSP is defined as follows.

Definition 8. *Let $\mathcal{P} := \{\mathcal{P}_i\}_{i=1}^K$ be a partition of the index set $\{1,\cdots,N\}$ and* $\text{Null}(\mathbf{A})$ *be the nullspace of* $\mathbf{A}$*. Denote a subset of* $\text{Null}(\mathbf{A})$ *w.r.t. (with respect to)* $\mathcal{P}$ *by*

$$\text{Null}(\mathbf{A},\mathcal{P}) := \{\boldsymbol{\eta}\in\text{Null}(\mathbf{A}) : \text{Supp}(\mathbf{A}) \nsubseteq P,\ \forall P\in\mathcal{P}\}.$$

The dictionary matrix $\mathbf{A}$ *satisfies the SNSP if and only if for any* $\eta\in\text{Null}(\mathbf{A},\mathcal{P})$ *and* $P\in\mathcal{P}$*, there holds*

$$\min_{\mathbf{z}:\mathbf{A}_P\mathbf{z}=\mathbf{A}_P\boldsymbol{\eta}_P}\|\mathbf{z}\|_1 < \|\boldsymbol{\eta}_{P^c}\|_1, \tag{31}$$

where P^c denotes the complement set of P.

Compared with SNSP, the proposed ARP in Definition 5 is more concise and intuitive. Furthermore, the work [Kaba *et al.*, 2021] is restricted to single-view data and cannot be applied for multi-view data. In comparison, our results can apply to both single-view and multi-view data.

Comparison to [Thaker *et al.*, 2022]. The work [Thaker *et al.*, 2022] considers the problem of reverse engineering adversarial attacks and poses it as a block sparse recovery (BSR) problem. 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 2 in our paper is *both necessary and sufficient* condition for multi-view subspace-preserving recovery while the core results in [Thaker *et al.*, 2022] are only sufficient for single-view subspace-preserving recovery. *Thirdly*, the results in

References	Year	Task	Model	Sufficient	Necessary	Multi-view
[Soltanolkotabi and Candés, 2012]	2012	Clustering	BP	✓	×	×
[Dyer *et al.*, 2013]	2013	Clustering	OMP	✓	×	×
[You and Vidal, 2015]	2015	Recovery	BP&OMP	✓	×	×
[You *et al.*, 2016]	2016	Clustering	OMP	✓	×	×
[Wang *et al.*, 2019]	2019	Classification	AR	✓	×	×
[Kaba *et al.*, 2021]	2021	Recovery	BP	✓	✓	×
[Thaker *et al.*, 2022]	2022	Recovery	BSP	✓	×	×
This paper	2024	Recovery	JSR	✓	✓	✓

Table 2: Comparison of different related works. Here "Sufficient" and "Necessary" mean whether the theoretical conditions in the corresponding works are sufficient and necessary, respectively.

[Thaker *et al.*, 2022] requires the data labels in advance while our results do not depend on such supervised information.

Based on the analysis above, we can draw the conclusion that our theoretical results for MSPR are novel and meaningful compared to prior works.

4 Application to Multi-view Subspace Clustering

Consider the Multi-view Sparse Subspace Clustering (MSSC) model [Wang *et al.*, 2023]

$$\begin{aligned}\min_{\mathbf{C}_i\in\mathbb{R}^{N\times V}} \|\mathbf{C}_i\|_{1,2}, \text{ s.t. } \mathbf{x}_i^v = \mathbf{X}^v\mathbf{C}_i(:,v), \\ \mathbf{C}_i(i,:) = \mathbf{0}, \ v\in[V],\end{aligned} \tag{32}$$

where $\mathbf{C}_i(i,:)$ is the i-th row of the matrix $\mathbf{C}_i$. Let $\mathbf{C}_i^*$ be the optimal solution of (32) for $i = 1,\cdots,N$. We first construct the common representation matrix $\mathbf{C}^*\in\mathbb{R}^{N\times N}$ such that

$$\mathbf{C}^*(:,i) = \left[\|\mathbf{C}_i^*(1,:)\|_2, \cdots, \|\mathbf{C}_i^*(N,:)\|_2\right]^T, \tag{33}$$

for $i = 1,\cdots,N$. Then we construct the affinity matrix $\mathbf{W} = |\mathbf{C}^*| + \left|(\mathbf{C}^*)^T\right|$. Finally, we apply the spectral clustering algorithm to the affinity matrix and obtain the final clustering results. A desirable affinity matrix $\mathbf{W}$ should satisfy the following subspace-preserving property: $W_{ij}\neq 0$ only if the i-th sample and the j-th sample are from the same cluster.

Theorem 4. *The affinity matrix $\mathbf{W}$ generated by the MSSC model (32) satisfies the subspace-preserving property if and only if the leave-one-out multi-view data $\dot{\mathbf{X}}\backslash\dot{\mathbf{x}}_i$ satisfies the ARP for $i = 1,\cdots,N$.*

Proof. Assume that $\dot{\mathbf{X}}\backslash\dot{\mathbf{x}}_i$ satisfies the ARP. According to Theorem 1, for each sample $\{\mathbf{x}_i^v\}_{v=1}^V$, the representation matrix $\mathbf{C}_i^*$ is a MSPR, i.e., supprow$(\mathbf{C}_i^*)\subset\mathcal{I}_k$. It follows from Eq. (33) that each column $\mathbf{C}^*(:,i)$ of $\mathbf{C}^*$ is also subspace preserving, i.e., $\mathbf{C}^*(j,i)\neq 0$ only if $\mathbf{x}_i^v$ and $\mathbf{x}_j^v$ are from the same subspace for each view. According to the construction of the affinity matrix $\mathbf{W}$, $W_{ij}\neq 0$ only if i-th sample and the j-th sample are from the same cluster. Thus, $\mathbf{W}$ satisfies the subspace-preserving property and completes the proof. ☐

5 Application to Multi-view Subspace Classification

Given any new multi-view test sample $\dot{\mathbf{y}}\in\dot{\mathcal{S}}_k\backslash\{\dot{\mathbf{0}}\}$, the MSC (Multi-view Subspace Classification) method [Shekhar *et al.*, 2013] first computes the representation matrix of $\dot{\mathbf{y}}$ using the following JSR model

$$\min_{\mathbf{C}\in\mathbb{R}^{N\times V}} \|\mathbf{C}\|_{1,2}, \text{ s.t. } \mathbf{y}^v = \mathbf{D}^v\mathbf{C}(:,v), \ \forall v\in[V], \tag{34}$$

where $\dot{\mathbf{D}}$ denotes the multi-view training data. Then for each class the reconstruction residual is calculated as

$$r_k(\dot{\mathbf{y}}) = \sum_{v=1}^V \|\mathbf{y}^v - \mathbf{D}^v\delta_k(\mathbf{C}(:,v))\|_2, \ k = 1,\cdots,K \tag{35}$$

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 5. *MSC 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 $\dot{\mathbf{D}}$ satisfy the ARP property.*

Proof. According to Theorem 1, the representation matrix $\mathbf{C}^*$ of the test sample $\dot{\mathbf{y}}\in\dot{\mathcal{S}}_k\backslash\{\dot{\mathbf{0}}\}$ is a MSPR, i.e., $\mathbf{C}^*(\mathcal{I}_{-k},:) = \mathbf{0}$. It follows that

$$\mathbf{y}^v = \mathbf{D}^v\mathbf{C}(:,v) = \mathbf{D}^v\delta_k(\mathbf{C}(:,v)), \ \forall v\in[V].$$

According to Eq. (35), the reconstruction residual of each class is $r_k(\dot{\mathbf{y}}) = \sum_{v=1}^V\|\mathbf{y}^v - \mathbf{D}^v\delta_k(\mathbf{C}(:,v))\|_2 = 0$, and $r_l(\dot{\mathbf{y}}) = \sum_{v=1}^V\|\mathbf{y}^v - \mathbf{D}^v\delta_l(\mathbf{C}(:,v))\|_2 = \sum_{v=1}^V\|\mathbf{y}^v\|_2$, $l\neq k$. Since $r_k(\dot{\mathbf{y}}) < \min_{l\neq k} r_l(\dot{\mathbf{y}})$, the test sample is assigned to the correct class k. ☐

6 Conclusions

This paper studies the multi-view subspace-preserving recovery theory, which plays a critical role for multi-view subspace clustering and classification. We first generalize the atomic norm to multi-view data and define the Multi-view Atomic Norm (MAN). Based on MAN, we derived a necessary and sufficient condition dubbed ARP (Atomic Recovery Property) for the JSR model to produce MSPR. To the best of our knowledge, ARP is the first necessary and sufficient theoretical condition for multi-view subspace-preserving recovery. The results reveal important geometric sights and provide theoretical justification for the success of multi-view subspace clustering and classification.

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. 62076041, 62276111, 61702057, 12071166, 62376104 and 62172458), the Fundamental Research Funds for the Central Universities of China (under Grant Nos. 2662023XXPY002 and 2662023LXPY005), the HZAU-AGIS Cooperation Fund (under Grant No. SZYJY2023010), and the Industry-University-Research Cooperation Project of Zhuhai, Guangdong, China, under Grant GP/026/2020 and Grant HF-010-2021.

References

[Abavisani and Patel, 2018] Mahdi Abavisani and Vishal M Patel. Deep multimodal subspace clustering networks. *IEEE J. Sel. Top. Signal Process.*, 12(6):1601–1614, 2018.

[Bahrampour *et al.*, 2015] Soheil Bahrampour, Nasser M Nasrabadi, Asok Ray, and William Kenneth Jenkins. Multimodal task-driven dictionary learning for image classification. *IEEE Trans. Image Process.*, 25(1):24–38, 2015.

[Chandrasekaran *et al.*, 2012] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

[Chen *et al.*, 1998] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33, 1998.

[Chen *et al.*, 2023] Zhe Chen, Xiao-Jun Wu, Tianyang Xu, and Josef Kittler. Fast self-guided multi-view subspace clustering. *IEEE Trans. Image Process.*, 32:6514–6525, 2023.

[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.

[Foucart and Rauhut, 2013] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer Science & Business Media, 2013.

[Guo *et al.*, 2022] Jipeng Guo, Yanfeng Sun, Junbin Gao, Yongli Hu, and Baocai Yin. Logarithmic schatten-p norm minimization for tensorial multi-view subspace clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(3):3396–3410, 2022.

[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.*, 2019] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. Flexible multi-view representation learning for subspace clustering. In *IJCAI*, pages 2916–2922, 2019.

[Pati *et al.*, 1993] Yagyensh Chandra Pati, Ramin Rezaiifar, and Perinkulam Sambamurthy Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar conference on signals, systems and computers*, pages 40–44 vol.1, Nov. 1993.

[Ren *et al.*, 2022] Hengyi Ren, Lijuan Sun, Jian Guo, and Chong Han. A dataset and benchmark for multimodal biometric recognition based on fingerprint and finger vein. *IEEE Trans. Inf. Forensics Secur.*, 17:2030–2043, 2022.

[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.

[Tang and Liu, 2022] Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, pages 21090–21110, 2022.

[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.

[Tsakiris and Vidal, 2018] Manolis Tsakiris and Rene Vidal. Theoretical analysis of sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pages 4975–4984, 2018.

[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.

[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.

[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.

[Zhang *et al.*, 2017] Heng Zhang, Vishal M Patel, and Rama Chellappa. Low-rank and joint sparse representations for multi-modal recognition. *IEEE Transactions on Image Processing*, 26(10):4741–4752, 2017.

[Zhang *et al.*, 2021] Shangzhi Zhang, Chong You, René Vidal, and Chun-Guang Li. Learning a self-expressive network for subspace clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12393–12403, 2021.