Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.OA-2025-0081

Vol. **xx**, No. x, pp. 1-28
x 2026

Sequential Subspace Clustering via Joint Capped ℓ_2 and $\ell_{2,p}$ Norm Minimization with Convergence Guarantee

Zhihui Tu[1], Jian Lu[2], Wenyu Hu[3,4,*] and Shan Liu[1]

[1] *School of Mathematical Sciences, Jiangxi Science and Technology Normal University, Nanchang 330038, China*
[2] *School of Mathematical Sciences, Shenzhen University, Shenzhen 518060, China*
[3] *School of Mathematics and Computer Sciences, Gannan Normal University, Ganzhou 341000, China*
[4] *Key Laboratory of Data Science and Artificial Intelligence of Jiangxi Education Institutes, Ganzhou 341000, China*

Received 13 August 2025; Accepted (in revised version) 4 January 2026

Abstract. Subspace clustering is a fundamental problem in machine learning that has attracted considerable attention in recent years. Most existing methods focus on designing effective models to regularize the coefficient matrix, often neglecting the impact of noise on subspace structures. However, real-world data are typically corrupted by noise, which can distort the underlying subspace structure. Additionally, for sequential data, a key challenge is the effective exploitation of temporal information. To address these issues, we propose a novel and robust sequential subspace clustering method, termed joint capped ℓ_2 and $\ell_{2,p}$ norm minimization (JCLLM). The capped ℓ_2 norm-based loss function mitigates the influence of noise and outliers in regression, while the $\ell_{2,p}$ norm regularization captures the temporal dependencies inherent in sequential data. We develop an iteratively reweighted optimization algorithm to solve the JCLLM model and prove its convergence to a stationary point. Extensive experiments on both synthetic and real-world datasets demonstrate that our method consistently outperforms several state-of-the-art subspace clustering approaches.

AMS subject classifications: 62H30, 65F35, 68U10

Key words: Sequential data, capped ℓ_2 norm, $\ell_{2,p}$ norm, subspace clustering.

1. Introduction

In recent years, subspace clustering has emerged as a significant research topic in computer vision due to its strong performance in various real-world applications, such

*Corresponding author. *Email address:* cswenyuhu@163.com (W. Hu)

as face clustering [7,9,11,26,40], handwritten digit clustering [17], motion segmentation [19–21], and video segmentation [25, 37, 39]. Given a set of data points sampled from a union of subspaces, the goal of subspace clustering is to determine the number of subspaces, their dimensions, the basis for each subspace, and the corresponding data segmentation. The core idea is to learn an effective representation that facilitates the construction of an affinity matrix, upon which spectral clustering is applied to partition the data [35].

Over the past two decades, numerous subspace clustering methods have been proposed, including iterative methods [3], algebraic methods [8, 39], statistical methods [32], and spectral-based methods [10,27]. Among these, spectral clustering-based methods have demonstrated superior performance on real-world data. In this paper, we focus exclusively on spectral clustering-based approaches. These methods typically involve two main steps: constructing an affinity matrix and applying a clustering algorithm based on the affinity matrix to partition the data. Clearly, constructing an effective affinity matrix is a crucial step in spectral clustering-based methods. The self-expression model is commonly employed for this purpose. It treats the original data as a dictionary and can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E}} \quad & \phi(\mathbf{E}) + \lambda\delta(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{Z} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \quad \text{diag}(\mathbf{Z}) = 0, \end{aligned} \tag{1.1}$$

where $\mathbf{X}$ denotes the data matrix, where each column represents a sample point; $\mathbf{Z}$ is the coding matrix, and $\mathbf{E}$ is the error matrix. The functions $\phi(\cdot)$ and $\delta(\cdot)$ are designed to act on $\mathbf{E}$ and $\mathbf{Z}$, respectively, to enforce different types of prior knowledge inherent in the data. The parameter $\lambda > 0$ serves as a trade-off between the reconstruction error and the regularization terms. Additionally, the constraint $\text{diag}(\mathbf{Z}) = 0$ is optionally applied to avoid the trivial solution where $\mathbf{Z}$ becomes an identity matrix.

The primary differences among most subspace clustering methods lie in the choice of the regularization function $\delta(\mathbf{Z})$ applied to the representation matrix $\mathbf{Z}$, which serves to enhance the global structural information of the data. We refer to $\delta(\mathbf{Z})$ as the global structure regularization function. Representative methods include sparse subspace clustering (SSC) [10], low-rank representation (LRR) [27], and least squares regression (LSR) [30]. Specifically, SSC employs the ℓ_1 norm as $\delta(\mathbf{Z})$ to obtain the sparsest representation for each data point; LRR uses the nuclear norm $\|\cdot\|_*$ as a convex surrogate for the rank function to yield a low-rank coefficient matrix; and LSR constrains the representation matrix $\mathbf{Z}$ with the Frobenius norm to achieve a block-diagonal structure. Moreover, to enhance the performance of SSC and LRR, several variant methods have been proposed [9, 26, 28, 41, 42]. For instance, to address limitations posed by small-scale datasets, Liu *et al.* [28] introduced the latent low-rank representation (LatLRR) method, which constructs the dictionary using both explicit observed data and implicit latent data, effectively overcoming the performance bottleneck of LRR. To further explore the intrinsic structure of data, Ding *et al.* [9] proposed a bilateral tensor low-rank representation (BTLRR) method, which fully utilizes the bi-

lateral information of observed data, capturing not only the similarities among samples but also the correlations among features. However, although methods based on SSC and LRR and their extensions have achieved remarkable success in conventional data processing, they still struggle to fully exploit the inherent information contained in data with special characteristics, such as the temporal priors inherent in sequential data.

Sequential data, such as videos, are ubiquitous in real-world applications [7,22,40]. A key characteristic of such data is that adjacent frames often share similar features and thus may lie in the same subspace. This observation motivates the development of unsupervised visual learning algorithms tailored for sequential data, a task known as sequential subspace clustering. In general, to perform sequential subspace clustering, a specific regularization term $f(\mathbf{Z})$ is incorporated into the self-expressive model (1.1), which can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E}} \quad & \phi(\mathbf{E}) + \lambda_1 \delta(\mathbf{Z}) + \lambda_2 f(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \quad \text{diag}(\mathbf{Z}) = 0, \end{aligned} \tag{1.2}$$

where $f(\mathbf{Z})$ represents local or temporal information used as a regularization term on $\mathbf{Z}$ to enforce the temporal properties of sequential data. The parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ are trade-off coefficients. We refer to $f(\mathbf{Z})$ as local temporal regularization. Guo *et al.* [16] proposed a spatial subspace clustering method (SpatSC) by introducing an additional ℓ_1-norm temporal term into the SSC model to exploit the temporal information in the data. However, the ℓ_1-norm can only ensure similarity between adjacent elements in the matrix and fails to effectively capture the overall structural similarity between neighboring columns. To overcome this limitation, Tierney *et al.* [37] replaced the ℓ_1-norm with the $\ell_{2,1}$-norm and proposed the ordered subspace clustering (OSC) method, which better models the global temporal relationships in the data. However, OSC neglects the local structural information between data points. To address this issue, Li *et al.* [25] incorporated Laplacian regularization and proposed the temporal subspace clustering (TSC) method, which was successfully applied to human motion segmentation tasks with promising results. Furthermore, to explore the intrinsic relationship between the structural regularization term $\delta(\mathbf{Z})$ and the local temporal regularization term $f(\mathbf{Z})$ more deeply, Liu *et al.* [29] proposed the temporal smoothness sequential subspace clustering (TS^3C) method based on the idea of temporal predictability. This method demonstrates that, in model design, constructing $f(\mathbf{Z})$ appropriately is more critical than $\delta(\mathbf{Z})$. Some recent studies have investigated robust multi-view clustering, such as hypergraph-regularized deep non-negative matrix factorization [6] and robust weighted low-rank tensor approximation under mixed noise [34].

However, real-world data are often contaminated by noise [31], and the statistical distribution of this noise is typically complex. As noted in [24, 27], if the noise cannot be effectively characterized, the coefficient matrix will fail to capture the relationships among data points, thereby degrading the performance of subspace clustering. Most existing subspace clustering methods, particularly those that rely on squared error functions applied to observed entries, are sensitive to both noise and outliers.

To address the aforementioned problem, we propose a novel and efficient sequential subspace clustering method, termed joint capped ℓ_2 norm loss function and $\ell_{2,p}$ norm local temporal regularization minimization. Specifically, we employ a capped ℓ_2 norm-based loss function to mitigate the influence of outliers, thereby enhancing robustness to noise. The capped ℓ_2 norm is defined as follows [23]:

$$\|\mathbf{X}\|_{C_2^\varepsilon} = \sum_i \min\left(\|\mathbf{x}_i\|_2, \varepsilon\right),$$

where $\mathbf{x}_i$ denotes the i-th sample point in the data matrix $\mathbf{X}$, and ε is a threshold parameter that controls the upper bound of the norm. Compared to the conventional ℓ_2 norm, the capped ℓ_2 norm remains constant when $\|\mathbf{x}_i\|_2$ exceeds ε, as illustrated in Fig. 1. This implies that noise with large magnitudes can be effectively suppressed. Consequently, the capped ℓ_2 norm exhibits greater robustness to noise than the traditional ℓ_2 norm. Furthermore, extensive computational studies [4, 5, 12] have demonstrated that employing the ℓ_p norm ($0 < p < 1$) can yield sparser solutions than the ℓ_1 norm. Naturally, minimization based on the mixed $\ell_{2,p}$ norm ($0 < p \leq 1$) is expected to produce superior sparsity patterns compared to the $\ell_{2,1}$ norm. Unlike traditional methods such as SpatSC and OSC that utilize the ℓ_1 or $\ell_{2,1}$ norm for spectral embedding, our model incorporates a $\ell_{2,p}$ norm-based temporal regularization term, making it more resilient to noise and outliers in sequential data. Last but not least, our proposed objective jointly minimizes the capped ℓ_2 norm and the $\ell_{2,p}$ norm, yielding a model that is both more robust and more efficient than standard subspace clustering approaches for sequential data. Although the resulting optimization problem is non-convex, we develop an efficient iterative algorithm with rigorous convergence analysis. Extensive experiments on both synthetic and real-world datasets demonstrate the superiority of our method over several state-of-the-art clustering algorithms. In summary, we high-

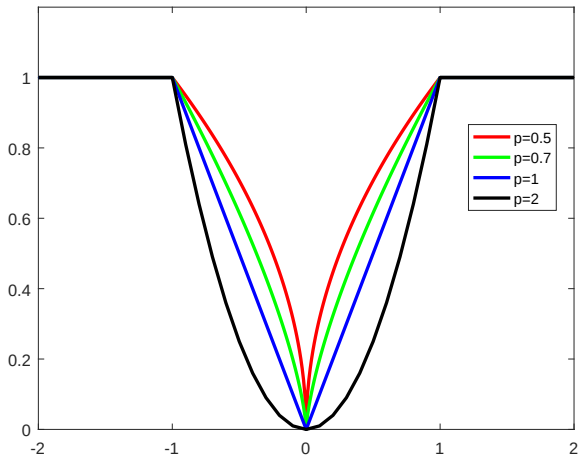


Figure 1: An example of capped ℓ_p norm (ε=1).

light the main contributions of this paper as follows:

- We propose an efficient method for robust sequential subspace clustering by jointly employing the capped ℓ_2 norm and the $\ell_{2,p}$ norm. This approach enhances robustness to outliers, enforces joint sparsity, and effectively captures local temporal structures.
- An alternating iterative algorithm is developed to solve the resulting optimization problem. We also provide a convergence analysis and a detailed parameter study of the proposed method.
- Extensive experiments are conducted to evaluate the performance of our algorithm, demonstrating that it consistently outperforms existing methods.

The remainder of this paper is organized as follows. Section 2 provides a brief review of the SpatSC, OSC, TSC, and TS^{3}C models. In Section 3, we introduce the proposed JCLM method, which is based on the capped ℓ_2 norm and the $\ell_{2,p}$ norm. Extensive experiments are presented in Section 4. Finally, Section 5 concludes the paper.

2. A general framework of sequential subspace clustering

2.1. Notation

In this paper, matrices and vectors are denoted by bold uppercase and bold lowercase letters, respectively. For example, let $\mathbf{M} \in \mathbb{R}^{m\times n}$ be a matrix; $\mathbf{M}_{ij}$ denotes its (i,j)-th entry, $\mathbf{m}^i$ denotes its i-th row, and $\mathbf{m}_j$ denotes its j-th column. The Frobenius norm of the matrix $\mathbf{M}$ is defined as

$$\|\mathbf{M}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} m_{ij}^2}.$$

For matrix $\mathbf{M}$, the $\ell_{2,1}$ norm is defined as $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{n} \|\mathbf{m}_i\|_2$. This can be generalized to the $\ell_{2,p}$ norm

$$\|\mathbf{M}\|_{2,p} = \left(\sum_{i=1}^{n} \|\mathbf{m}_i\|_2^p\right)^{\frac{1}{p}}.$$

When $p \geq 1$, the $\ell_{2,p}$ norm satisfies the properties of a valid norm, including the triangle inequality

$$\|\mathbf{A}\|_{2,p} + \|\mathbf{B}\|_{2,p} \geq \|\mathbf{A}+\mathbf{B}\|_{2,p}.$$

However, when $0 \leq p < 1$, the $\ell_{2,p}$ norm (as well as the ℓ_0 norm) does not satisfy all norm axioms, and is thus not a true norm in the strict mathematical sense. Nevertheless, we still refer to them as norms for the sake of convenience.

2.2. Related work

Consider a sequential data matrix composed of column-wise samples, denoted as $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, where each sample is drawn from a union of k linear subspaces $\{S_i\}_{i=1}^k$. Here, d is the dimensionality of each sample, and n is the total number of samples. The objective of sequential subspace clustering is to segment the data according to the underlying subspaces from which the samples originate. In this section, we briefly review three state-of-the-art sequential subspace clustering methods: SpatSC [16], OSC [37], and TSC [25]. All of these methods incorporate an additional penalty term on the representation matrix $\mathbf{Z}$ into traditional subspace clustering frameworks, which are typically based on global structural regularization of $\mathbf{Z}$, in order to better exploit the sequential nature of the data.

1. SpatSC: SpatSC is a variant of the SSC method that introduces a local temporal regularization term on $\mathbf{Z}$ to directly constrain the similarity between consecutive columns while preserving the sparsity of $\mathbf{Z}$. This method encodes the sequential information using a matrix $\mathbf{R}$, which enforces similarity between adjacent columns of the coding matrix $\mathbf{Z}$ (i.e., $\mathbf{z}_i \approx \mathbf{z}_{i+1}$). As a result, the following optimization problem is formulated:

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{Z}\mathbf{R}\|_1 \\ \text{s.t.} \quad & \operatorname{diag}(\mathbf{Z}) = 0, \end{aligned} \tag{2.1}$$

   where $\mathbf{R}$ is a lower triangular matrix, to enforce the consistency of the spare representation of the sequential data

$$\mathbf{R} = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{n \times (n-1)}. \tag{2.2}$$

2. OSC: OSC method adopts the $\ell_{2,1}$ norm in the third term instead of ℓ_1 norm, and gets the following model:

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{Z}\mathbf{R}\|_{2,1} \\ \text{s.t.} \quad & \operatorname{diag}(\mathbf{Z}) = 0. \end{aligned} \tag{2.3}$$

   The penalty term $\|\mathbf{Z}\mathbf{R}\|_{2,1}$ is stronger than $\|\mathbf{Z}\mathbf{R}\|_1$ to enforce column similarity of $\mathbf{Z}$. However, the applied squared error is sensitive to outliers, and the ADMM method used to solve problem (2.3) in [37] is not guaranteed to converge.

3. TSC: TSC is a variant of LSR method, which adds a term of local temporal regularization on $\mathbf{Z}$ and learns a non-negative dictionary from data with a temporal Laplacian regularization, whose goal is to encode the temporal relationships in sequential data and obtain more expressive codings. Its model is

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{D}} \quad & \|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2+\lambda_1\|\mathbf{Z}\|_F^2+\lambda_2 f(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{Z}\geq 0, \quad \mathbf{D}\geq 0, \quad \|\mathbf{d}_i\|_2^2\leq 1, \quad i=1,2,\ldots,r, \end{aligned} \tag{2.4}$$

where the non-negative constraints $\mathbf{Z}\geq 0$ and $\mathbf{D}\geq 0$ are to ensure the learned bases and corresponding representation codings have non-negative values, the constraint $\|\mathbf{d}_i\|_2^2\leq 1$ to control the model complexity, and $f(\mathbf{Z})$ is the temporal Laplacian regularization function. Given a representation matrix $\mathbf{Z}$, it is defined as

$$f(\mathbf{Z})=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}\|\mathbf{z}_i-\mathbf{z}_j\|_2^2=\mathrm{tr}\big(\mathbf{Z}\mathbf{L}_\mathbf{T}\mathbf{Z}^T\big), \tag{2.5}$$

where $\mathbf{L}_T$ is the temporal Laplacian matrix, $\mathbf{L}_T=\widetilde{\mathbf{D}}-\mathbf{W}$, $\widetilde{\mathbf{D}}_{ii}=\sum_{j=1}^{n}\omega_{ij}$, $\mathbf{W}$ is the weight matrix that captures the sequential relationships in $\mathbf{X}$. Let s denote the number of sequential neighbors for each sample, the element in $\mathbf{W}$ is calculated as

$$\omega_{ii}=\begin{cases}1, & \text{if} \quad |i-j|\leq\dfrac{s}{2},\\ 0, & \text{otherwise.}\end{cases}$$

4. TS^{3}C: For sequential data, Liu *et al.* [29] design a novel local temporal regularization term on $\mathbf{Z}$ to enforce the temporal smoothness of sequential data, based on the concept of temporal predictability [36]. The resulted model is

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{D}} \quad & \|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2+\lambda\|\mathbf{Z}\mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{B}=\mathbf{I}-\mathbf{T}^\top, \quad \|\mathbf{d}_i\|_2^2\leq 1, \quad i=1,\ldots,r, \end{aligned} \tag{2.6}$$

where $\mathbf{B}=\mathbf{I}-\mathbf{T}^\top$ is an auxiliary variable and $\mathbf{T}$ is a $n\times n$ Toeplitz matrix with the right-shifted template appearing an each row, and $^\top$ denotes the matrix transposition

$$\mathbf{T}=\begin{pmatrix} \beta & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \alpha\beta & \beta & 0 & 0 & 0 & 0 & \cdots & 0 \\ \alpha^2\beta & \alpha\beta & \beta & 0 & 0 & 0 & \cdots & 0 \\ \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta & 0 & 0 & \cdots & 0 \\ 0 & \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta & 0 & \cdots & 0 \\ 0 & 0 & \cdots & & & & \ddots & \vdots \\ 0 & \cdots & & 0 & \alpha^3\beta & \alpha^2\beta & \alpha\beta & \beta \end{pmatrix}. \tag{2.7}$$

TS^{3}C demonstrates that a properly designed local temporal regularization term $f(\mathbf{Z})$ can be more effective than the global structural regularization term $\delta(\mathbf{Z})$. In

some cases, it may not even be necessary to include the global structural regularization. Motivated by this insight, we propose a sequential subspace clustering model that jointly minimizes a capped ℓ_2 norm loss and a $\ell_{2,p}$ norm-based regularization term, resulting in an objective function with only two components.

2.3. The general framework of sequential subspace clustering

Based on Eq. (1.2), we give the unified model of sequential subspace clustering methods as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E}} \quad & \phi(\mathbf{E}) + \lambda_1 \delta(\mathbf{Z}) + \lambda_2 f(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E}, \end{aligned} \tag{2.8}$$

which consists of three terms, the reconstruction error $\phi(\mathbf{E})$, the global structure regularization $\delta(\mathbf{Z})$ and the local temporal regularization $f(\mathbf{Z})$ with the signal representation constraint $\mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E}$, where $\mathbf{D}$ is the dictionary to be learned or the original data samples $\mathbf{X}$ (in self-expressive model).

The most representative existing methods, including SSC [10], LRR [27], LSR [30], SpatSC [16], OSC [37], TSC [25], and TS^{3}C [29], are analyzed under the framework of the unified model from the following aspects:

- **Model Components:** These methods incorporate a combination of global structure regularization $\delta(\mathbf{Z})$ and local temporal regularization $f(\mathbf{Z})$, as summarized in Table 1. Most of them utilize the Frobenius norm $\|\mathbf{E}\|_F$ to measure the reconstruction error. However, the Frobenius norm is known to be sensitive to outliers [33].

- **Optimization Techniques:** Since all existing subspace clustering models are formulated as constrained optimization problems, the alternating direction method of multipliers (ADMM) [2] is commonly employed to solve them. ADMM is a hybrid method that combines the augmented Lagrangian multiplier (ALM) technique and the alternating direction method (ADM) [38], and solves separable unconstrained subproblems using a block coordinate descent (or alternating minimization) approach. More details can be found in [2] and the references therein. SSC, LRR, OSC, and TSC adopt ADMM for optimization, whereas LSR has an analytical solution.

Table 1: The analysis of existing methods in the framework of unified model.

Methods	SSC	LRR	LSR	spatSC	OSC	TSC	TS^{3}C	JCLLM
$\phi(\mathbf{E})$	$\|\mathbf{E}\|_F^2$	$\|\mathbf{E}\|_{2,1}$	$\|\mathbf{E}\|_F^2$	$\|\mathbf{E}\|_F^2$	$\|\mathbf{E}\|_F^2$	$\|\mathbf{E}\|_F^2$	$\|\mathbf{E}\|_F^2$	$\|\mathbf{E}\|_{C_2^\varepsilon}$
$\delta(\mathbf{Z})$	$\|\mathbf{Z}\|_1$	$\|\mathbf{Z}\|_*$	$\|\mathbf{Z}\|_F$	$\|\mathbf{Z}\|_1$	$\|\mathbf{Z}\|_1$	$\|\mathbf{Z}\|_F$	n/a	n/a
$f(\mathbf{Z})$	n/a	n/a	n/a	$\|\mathbf{Z}\mathbf{R}\|_1$	$\|\mathbf{Z}\mathbf{R}\|_{2,1}$	$tr(\mathbf{Z}\mathbf{L}_T\mathbf{Z}^T)$	$\|\mathbf{Z}\mathbf{B}\|_F^2$	$\|\mathbf{Z}\mathbf{R}\|_{2,p}$

3. Joint capped ℓ_2 norm and $\ell_{2,p}$ norm minimization for sequential subspace clustering

In this section, we will introduce the formulation of our method and then propose a program to solve its objective function.

3.1. Formulation of JCLLM

In real-world applications, outliers frequently occur in datasets; therefore, traditional loss functions may fail to achieve optimal performance. A straightforward approach to address outliers is to detect and mitigate their impact. As illustrated in Fig. 1, the value of the capped ℓ_2 norm does not continuously increase with the magnitude of the error but instead becomes constant when $\|\mathbf{x}_i\|_2$ exceeds a predefined threshold ε. This property effectively reduces the influence of large outliers and noise in subspace clustering. Owing to its robustness to outliers and noise, we adopt a capped ℓ_2 norm-based loss function to penalize the noise term in our model

$$\min_{\mathbf{Z}} \sum_{i=1}^{n} \min\left(\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2, \varepsilon\right). \tag{3.1}$$

If the k-th element of $\mathbf{X}$ contains outliers, $\mathbf{x}_k$ will deviate significantly from the regression model. Consequently, the value of $\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2$ may become excessively large. To mitigate the influence of outliers, we employ the capped norm along with a threshold parameter ε. In this context, the selection and analysis of the threshold parameter ε are of critical importance. Through rigorous theoretical analysis and empirical evaluations, we conclude that the parameter ε indeed affects the final results, and our method remains robust as long as ε varies within a certain range.

Then, the next step is to add a regularization term $f(\mathbf{Z})$ with parameter γ to enforce the temporal properties of the sequential data. Based on the analysis of Eqs. (1.1) and (1.2), whose main difference lies in the two regularization terms — global structural regularization $\delta(\mathbf{Z})$ and local temporal regularization $f(\mathbf{Z})$ — we argue that, for specific tasks such as sequential subspace clustering, a properly designed $f(\mathbf{Z})$ is more important than $\delta(\mathbf{Z})$ [29]. In some cases, the inclusion of $\delta(\mathbf{Z})$ may even be unnecessary. Furthermore, we adopt the p-th power of the $\ell_{2,p}$ norm to capture the local temporal structure. The rationale is that the regularization term $\|\mathbf{Z}\mathbf{R}\|_{2,p}^p$ achieves similar results to $\|\mathbf{Z}\mathbf{R}\|_{2,p}$ under the regulation of γ, while enabling a more efficient optimization approach for solving the formulation.

In conclusion, we formulate our method as a joint minimization problem involving the capped ℓ_2 norm and the $\ell_{2,p}$ norm, as follows:

$$\min_{\mathbf{Z}} \sum_{i=1}^{n} \min(\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2, \varepsilon) + \gamma\|\mathbf{Z}\mathbf{R}\|_{2,p}^p. \tag{3.2}$$

Although the proposed model does not explicitly introduce a global structural regularization term $\delta(\mathbf{Z})$, the affinity structure is not degraded in practice for the follow-

ing reasons. First, the self-expressive reconstruction constraint inherently enforces the subspace-preserving property, since each data point is represented as a linear combination of other samples, and the dominant structural information is governed by the intrinsic data correlation contained in $\mathbf{X}^\top\mathbf{X}$. This mechanism naturally discourages cross-subspace connections even without imposing an explicit block-diagonal regularizer. Second, the capped ℓ_2 loss stabilizes the influence of large residuals by truncating extreme reconstruction errors. As a result, potential mis-assignments caused by noise, outliers, or abrupt transitions across different subspaces are effectively suppressed, preventing them from dominating the learned affinity matrix. Third, although the $\ell_{2,p}$ temporal regularizer encourages local consistency along the sequence, it does not blindly enforce smoothness. When two neighboring samples belong to different subspaces, imposing $\mathbf{z}_i \approx \mathbf{z}_{i+1}$ would induce a large reconstruction penalty. In such cases, the capped loss automatically weakens this erroneous temporal coupling, thereby avoiding incorrect affinity propagation across true subspace boundaries. Therefore, even without an explicit $\delta(\mathbf{Z})$ term, the combined effects of the self-expressive model, capped robust loss, and temporal regularization jointly preserve the block-diagonal affinity structure in practice.

Remark 3.1. Although the $\ell_{2,p}$ regularizer does not explicitly enforce sparsity on $\mathbf{Z}$, it does not lead to an over-connected affinity matrix in our model. This is because the block-diagonal structure is jointly preserved by the self-expressive reconstruction constraint and the capped ℓ_2 loss. Specifically, incorrect cross-subspace connections usually result in large reconstruction errors, whose influence is automatically suppressed by the capped loss. As a result, the affinity matrix remains block-diagonal in practice even without an explicit sparsity constraint.

3.2. Optimization and solution

In this section, we present an efficient method to solve the proposed formulation. Firstly, we denote diagonal matrix $\mathbf{H} \in \mathbb{R}^{n\times n}$ and the $i-$th diagonal element of $\mathbf{H}$ is

$$\mathbf{H}_{ii} = \begin{cases} \dfrac{1}{2\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2}, & \text{if} \quad \|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2 \leq \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

Then the formulation in Eq. (3.2) will be converted into the following problem:

$$\min_{\mathbf{Z}} \sum_{i=1}^{n} \mathbf{H}_{ii}\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2^2 + \gamma\|\mathbf{Z}\mathbf{R}\|_{2,p}^p. \tag{3.4}$$

As $\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2^2$ is equivalent to $(\mathbf{x}_i - \mathbf{X}\mathbf{z}_i)^\top(\mathbf{x}_i - \mathbf{X}\mathbf{z}_i)$. We can expand Eq. (3.4) to the following formulation:

$$\min_{\mathbf{Z}} \sum_{i=1}^{n} \mathbf{H}_{ii}(\mathbf{x}_i - \mathbf{X}\mathbf{z}_i)^\top(\mathbf{x}_i - \mathbf{X}\mathbf{z}_i) + \gamma\|\mathbf{Z}\mathbf{R}\|_{2,p}^p. \tag{3.5}$$

According to linear algebra and let $\mathbf{M} = \mathbf{ZR}$, we can reformulate Eq. (3.5) in matrix form as follows:

$$\mathcal{L}(\mathbf{Z}) = \mathrm{Tr}\left[(\mathbf{X} - \mathbf{XZ})\mathbf{H}(\mathbf{X} - \mathbf{XZ})^\top\right] + \gamma\|\mathbf{M}\|_{2,p}^p. \tag{3.6}$$

If $\mathbf{M}_i \neq 0$, $\|\mathbf{M}\|_{2,p}^p = \mathrm{Tr}(\mathbf{MDM}^\top)$, where

$$\mathbf{D}_{ii} = \frac{1}{2\|\mathbf{M}_i\|_p^{2-p}}. \tag{3.7}$$

Then the derivative of $\mathcal{L}$ in Eq. (3.6) can be regarded as the derivative of the following objective function:

$$\mathcal{L}(\mathbf{Z}) = \mathrm{Tr}\left[(\mathbf{X} - \mathbf{XZ})\mathbf{H}(\mathbf{X} - \mathbf{XZ})^\top\right] + \gamma\,\mathrm{Tr}\left(\mathbf{MDM}^\top\right). \tag{3.8}$$

That is to say,

$$\mathcal{L}(\mathbf{Z}) = \mathrm{Tr}\left[(\mathbf{X} - \mathbf{XZ})\mathbf{H}(\mathbf{X} - \mathbf{XZ})^\top\right] + \gamma\,\mathrm{Tr}\left(\mathbf{ZRDR}^\top\mathbf{Z}^\top\right). \tag{3.9}$$

Now, we adopt the objective function in Eq. (3.9) to approximate the JCLLM formulation presented in Eq. (3.2). Since both $\mathbf{H}$ and $\mathbf{D}$ depend on $\mathbf{Z}$, Eq. (3.9) becomes a regularized least squares problem when $\mathbf{H}$ and $\mathbf{D}$ are fixed. Therefore, we take the derivative of $\mathcal{L}(\mathbf{Z})$ with respect to $\mathbf{Z}$ and set it to zero, which leads to the following update rule for $\mathbf{Z}$:

$$\frac{\partial\mathcal{L}(\mathbf{Z})}{\partial\mathbf{Z}} = -\mathbf{X}^\top\mathbf{XH} + \mathbf{X}^\top\mathbf{XZH} + \gamma p\mathbf{ZRDR}^\top = 0, \tag{3.10}$$

or equivalently,

$$\mathbf{X}^\top\mathbf{XZ} + \gamma p\mathbf{ZRDR}^\top\mathbf{H}^{-1} = \mathbf{X}^\top\mathbf{X}. \tag{3.11}$$

Eq. (3.11) is the well-known Sylvester equation, which cost $\mathcal{O}(n^3)$ for a general solver. But if $\mathbf{X}^\top\mathbf{X}$ has certain structure, the costs may likely be $\mathcal{O}(n^2)$ [1]. We use the Matlab command `lyap` to solve Eq. (3.11) in this work.

We update $\mathbf{H}$ and $\mathbf{D}$ using Eqs. (3.3) and (3.7), respectively, based on the computed $\mathbf{Z}$. The matrix $\mathbf{Z}$ is considered an optimal solution to the problem if Eq. (3.11) is satisfied. In summary, we propose an iterative algorithm in this paper to obtain the solution $\mathbf{Z}$ such that Eq. (3.2) holds. In each iteration, $\mathbf{Z}$ is computed using the current $\mathbf{H}$ and $\mathbf{D}$, after which $\mathbf{H}$ and $\mathbf{D}$ are updated based on the newly computed $\mathbf{Z}$. The iterative procedure is repeated until the algorithm converges. Moreover, the experimental results on the convergence behavior demonstrate that the proposed algorithm converges rapidly. The main steps of our method are outlined in Algorithm 3.1.

After obtaining the representation matrix $\mathbf{Z}$ using Algorithm 3.1, the affinity graph is constructed as $(|\mathbf{Z}| + |\mathbf{Z}^\top|)/2$. Finally, the normalized cuts (Ncut) method [35] is employed to obtain the clustering results. Algorithm 3.2 summarizes the complete subspace clustering procedure.

Algorithm 3.1 Efficient Algorithm for Solving the Problem Eq. (3.2)

Input: Date matrix $\mathbf{X}$, parameters γ.
Initialize : $\mathbf{H}^0 = I$, $\mathbf{D}^0 = I$.
Output: Final coefficient matrix $\mathbf{Z}^*$.

1: **repeat**
2: Update $\mathbf{Z}^{(k+1)}$ by solving the following problem:

$$\mathbf{X}^T\mathbf{X}\mathbf{Z} + \gamma p\mathbf{Z}\mathbf{R}\mathbf{D}\mathbf{R}^\top\mathbf{H}^{-1} - \mathbf{X}^\top\mathbf{X} = 0.$$

3: Update $\mathbf{D}^{(k+1)}$

$$\mathbf{D}_{ii} = \frac{1}{2\|\mathbf{M}_i\|_p^{2-p}}.$$

4: Update $\mathbf{H}^{(k+1)}$

$$\mathbf{H}_{ii} = \begin{cases} \dfrac{1}{2\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2}, & \text{if} \quad \|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2 \leq \varepsilon, \\ 0, & \text{otherwise}. \end{cases}$$

5: **until** Converge

Algorithm 3.2 JCLLM

Input: Sequential data $\mathbf{X} \in \mathbb{R}^{d\times n}$, drawn from a union of k linear subspaces.

1: Obtain the optimal coefficient representation $\mathbf{Z}$ of $\mathbf{X}$ by Algorithm 3.1.
2: Create affinity matrix $\mathbf{A}$ using $\mathbf{Z}$, i.e., $\mathbf{A} = (|\mathbf{Z}| + |\mathbf{Z}^\top|)/2$. Then, an undirected graph can be obtained.
3: Apply Ncut to segment the vertices of the undirected graph into c clusters.

Output: The clustering results of $\mathbf{X}$.

Remark 3.2. In practice, to make the problem derivable and numerically stable, $\mathbf{H}_{ii} = 1/(2\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2)$ and $\mathbf{D}_{ii} = 1/(2\|\mathbf{M}_i\|_p^{2-p})$ are replaced by $\mathbf{H}_{ii} = 1/(2\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|_2 + \mu)$ and $\mathbf{D}_{ii} = 1/(2\|\mathbf{M}_i\|_p^{2-p} + \mu)$, respectively, where μ is a very small positive constant. This modification avoids division by zero and alleviates potential numerical instability caused by extremely small residuals or weights. It can be seen that when $\mu \to 0$, the modified formulation approximates the original problem.

3.3. Convergence and complexity analysis

Lemma 3.1. *Given*

$$s = \begin{cases} \dfrac{1}{2|e|}, & |e| < \varepsilon_1, \\ 0, & \textit{otherwise}, \end{cases}$$

we have the following inequality:

$$\min(|\widetilde{e}|,\varepsilon_1)-s\widetilde{e}^2\leq\min(|e|,\varepsilon_1)-se^2. \tag{3.12}$$

Proof. Let $h(x)=\min(\sqrt{e},\varepsilon)$, $\varepsilon>0$, notice $h(e)$ is concave function, and its derivative satisfy

$$h'(e)=\begin{cases}\dfrac{1}{2\sqrt{e}}, & 0<e<\varepsilon^2,\\ 0, & \text{otherwise.}\end{cases}$$

From the concave function property can be obtained,

$$\begin{aligned}&\min(|e|,\varepsilon_1)-\min(|\widetilde{e}|,\varepsilon_1)\\ &=h(e^2)-h(\widetilde{e}^2)\geq h'(e^2)(e^2-\widetilde{e}^2)\\ &=se^2-s\widetilde{e}^2.\end{aligned} \tag{3.13}$$

After sorting out the terms, Eq. (3.12) can be obtained. □

Lemma 3.2. *Assume each column of $\mathbf{X}\in\mathbb{R}^{m\times n}$ and $\mathbf{Y}\in\mathbb{R}^{m\times n}$ is nonzero. Let $g_i(x)$, $i=1,\ldots,n$, be concave and differentiable functions. We have*

$$\sum_{i=1}^{n}g_i\big(\|\mathbf{Y}_i\|_2^2\big)-g_i\big(\|\mathbf{X}_i\|_2^2\big)\geq\operatorname{Tr}\big((\mathbf{Y}^\top\mathbf{Y}-\mathbf{X}^\top\mathbf{X})\mathbf{D}\big), \tag{3.14}$$

where $\mathbf{D}\in\mathbb{R}^{n\times n}$ is a diagonal matrix, with its i-th diagonal element being $\mathbf{D}_{ii}=\nabla g_i(\|\mathbf{Y}_i\|_2^2)$.

By letting $g_i(x)=x^{q/2},0<q<2$, as a special case in Eq. (3.14) *we get*

$$\|\mathbf{Y}\|_{2,p}^p-\|\mathbf{X}\|_{2,p}^p\geq\frac{p}{2}\operatorname{Tr}\big((\mathbf{Y}^\top\mathbf{Y}-\mathbf{X}^\top\mathbf{X})\mathbf{D}\big). \tag{3.15}$$

Proof. By the definition of concave function, we have

$$\begin{aligned}&\sum_{i=1}^{n}g_i\big(\|\mathbf{Y}_i\|_2^2\big)-g_i\big(\|\mathbf{X}_i\|_2^2\big)\\ &\geq\sum_{i=1}^{n}\nabla g_i\big(\|\mathbf{Y}_i\|_2^2\big)\big(\|\mathbf{Y}_i\|_2^2-\|\mathbf{X}_i\|_2^2\big)\\ &=\operatorname{Tr}\big((\mathbf{Y}^\top\mathbf{Y}-\mathbf{X}^\top\mathbf{X})\mathbf{D}\big).\end{aligned}$$

The proof is complete. □

Theorem 3.1. *Let $\mathcal{J}(\mathbf{Z})$ denote the original objective function defined in Eq.* (3.2). *The sequence $\{\mathbf{Z}^{(k)}\}$ generated by Algorithm* 3.1 *satisfies the following properties:*

1. *The objective function $\mathcal{J}(\mathbf{Z})$ is non-increasing along the sequence $\{\mathbf{Z}^{(k)}\}$, i.e.,*

$$\mathcal{J}(\mathbf{Z}^{(k+1)}) \leq \mathcal{J}(\mathbf{Z}^{(k)}).$$

2. *The sequence $\{\mathbf{Z}^{(k)}\mathbf{R}\}$ is bounded.*

3. $\lim_{k\to\infty} \|\mathbf{Z}^{(k)}\mathbf{R} - \mathbf{Z}^{(k+1)}\mathbf{R}\|_F = 0$.

Proof. Since $\mathbf{Z}^{(k+1)}$ solve Eq. (3.10), we have

$$\mathbf{X}^T\big(\mathbf{X}\mathbf{Z}^{(k+1)} - \mathbf{X}\big)\mathbf{H}^{(k)} + \gamma p \mathbf{Z}^{(k+1)}\mathbf{R}\mathbf{D}^{(k)}\mathbf{R}^\top = 0. \tag{3.16}$$

According to Lemma 3.1, set $\widetilde{e} = \|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i^{(k+1)}\|_2$ and $e = \|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i^{(k)}\|_2$. Then we can have

$$\begin{aligned}
&\min\Big(\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i^{(k+1)}\|_2, \varepsilon\Big) - \mathbf{H}_{ii}^{(k)}\Big\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i^{(k+1)}\Big\|_2^2 \\
\leq &\min\Big(\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i^{(k)}\|_2, \varepsilon\Big) - \mathbf{H}_{ii}^{(k)}\Big\|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i^{(k)}\Big\|_2^2.
\end{aligned} \tag{3.17}$$

This together with Eq. (3.16) gives,

$$\begin{aligned}
&\|\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k)}\|_{C_2^\varepsilon} - \|\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k+1)}\|_{C_2^\varepsilon} \\
\geq\ &\mathbf{H}_{ii}^{(k)}\|\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k)}\|_2^2 - \mathbf{H}_{ii}^{(k)}\|\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k+1)}\|_2^2 \\
=\ &\operatorname{Tr}\big[(\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k)})^\top(\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k)}) - (\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k+1)})^\top(\mathbf{X} - \mathbf{X}\mathbf{Z}^{(k+1)})\mathbf{H}^{(k)}\big] \\
=\ &A + 2\operatorname{Tr}\big[(\mathbf{X}\mathbf{Z}^{(k)} - \mathbf{X}\mathbf{Z}^{(k+1)})^\top(\mathbf{X}\mathbf{Z}^{(k+1)} - \mathbf{X})\mathbf{H}^{(k)}\big] \\
=\ &A + 2\operatorname{Tr}\big[(\mathbf{Z}^{(k)} - \mathbf{Z}^{(k+1)})^\top\mathbf{X}^\top(\mathbf{X}\mathbf{Z}^{(k+1)} - \mathbf{X})\mathbf{H}^{(k)}\big] \\
=\ &A - 2\gamma p\operatorname{Tr}\big[(\mathbf{Z}^{(k)} - \mathbf{Z}^{(k+1)})^T\mathbf{Z}^{(k+1)}\mathbf{R}\mathbf{D}^{(k)}\mathbf{R}^\top\big] \\
=\ &A - 2\gamma p\operatorname{Tr}\big[\mathbf{R}^\top(\mathbf{Z}^{(k)} - \mathbf{Z}^{(k+1)})^\top\mathbf{Z}^{(k+1)}\mathbf{R}\mathbf{D}^{(k)}\big],
\end{aligned} \tag{3.18}$$

where

$$A = \operatorname{Tr}\big[(\mathbf{X}\mathbf{Z}^{(k)} - \mathbf{X}\mathbf{Z}^{(k+1)})^\top(\mathbf{X}\mathbf{Z}^{(k)} - \mathbf{X}\mathbf{Z}^{(k+1)})\mathbf{H}^{(k)}\big].$$

By using Eq. (3.15), we have

$$\begin{aligned}
&\gamma\|\mathbf{M}^{(k)}\|_{2,p}^p - \gamma\|\mathbf{M}^{(k+1)}\|_{2,p}^p \\
\geq\ &\gamma p\operatorname{Tr}\big[(\mathbf{M}^{(k)})^\top\mathbf{M}^{(k)} - (\mathbf{M}^{(k+1)})^\top\mathbf{M}^{(k+1)}\big]\mathbf{D}^{(k)} \\
=\ &B + 2\gamma p\operatorname{Tr}\big[(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})^\top(\mathbf{M}^{(k+1)} - \mathbf{M})\mathbf{D}^{(k)}\big] \\
=\ &B + 2\gamma p\operatorname{Tr}\big[(\mathbf{Z}^{(k)}\mathbf{R} - \mathbf{Z}^{(k+1)}\mathbf{R})^\top\mathbf{Z}^{(k+1)}\mathbf{R}\mathbf{D}^{(k)}\big] \\
=\ &B + 2\gamma p\operatorname{Tr}\big[\mathbf{R}^T(\mathbf{Z}^{(k)} - \mathbf{Z}^{(k+1)})^\top\mathbf{Z}^{(k+1)}\mathbf{R}\mathbf{D}^{(k)}\big],
\end{aligned} \tag{3.19}$$

where

$$B = \operatorname{Tr}\big[(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})^\top(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})\mathbf{D}^{(k)}\big].$$

Now, combining Eqs. (3.18) and (3.19) gives

$$\mathcal{J}(\mathbf{Z}^{(k)}) - \mathcal{J}(\mathbf{Z}^{(k+1)}) = A + B \geq 0. \tag{3.20}$$

The above equation implies the $\mathcal{J}(\mathbf{Z}^{(k)})$ is non-increasing. Then we have

$$\|\mathbf{Z}^{(k)}\mathbf{R}\|_{2,p}^{p} \leq \mathcal{J}(\mathbf{Z}^{(k)}) \leq \mathcal{J}(\mathbf{Z}^{(1)}) \triangleq N. \tag{3.21}$$

Thus, the sequence $\{\mathbf{Z}^{(k)}\}$ is bounded.

Furthermore,

$$\begin{aligned}
&\sum_{k=1}^{\infty} \gamma p \operatorname{Tr}\left[(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})^{\top}(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})\mathbf{D}^{(k)}\right] \\
\leq &\sum_{k=1}^{\infty} \left(\mathcal{J}(\mathbf{Z}^{(k)}) - \mathcal{J}(\mathbf{Z}^{(k+1)})\right) = \mathcal{J}(\mathbf{Z}^{(1)}) - \lim_{k\to\infty} \mathcal{J}(\mathbf{Z}^{(k)}) \leq \mathcal{J}(\mathbf{Z}^{(1)}).
\end{aligned}$$

According to the Cauchy criterion of series convergence, we have

$$\sum_{k=1}^{\infty} \operatorname{Tr}\left[(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})^{\top}(\mathbf{M}^{(k)} - \mathbf{M}^{(k+1)})\mathbf{D}^{(k)}\right] \to 0. \tag{3.22}$$

In particular, Eq. (3.22) implies that

$$\lim_{k\to\infty} \left\|\mathbf{Z}^{(k)}\mathbf{R} - \mathbf{Z}^{(k+1)}\mathbf{R}\right\|_F = 0.$$

The proof is complete. □

4. Experiments

In this section, we conduct numerical experiments on both synthetic and real datasets to demonstrate the effectiveness of the proposed JCLLM algorithm. We first examine the behavior of JCLLM and its sensitivity to the parameter μ, and then compare our method with several state-of-the-art subspace clustering algorithms.

In the experiment, we use clustering accuracy (Acc) to establish a one-to-one correspondence between clusters and ground-truth classes, and to evaluate the extent to which each cluster contains data points from its corresponding class. The Acc metric is defined as follows:

$$\mathrm{Acc} = \frac{1}{n}\sum_{i=1}^{n} \delta\big(p_i, \mathrm{map}(q_i)\big), \tag{4.1}$$

where p_i and q_i represent the output label and the ground truth one of the i-th point respectively, $\delta(x, y) = 1$ if $x = y$, and $\delta(x, y) = 0$ otherwise, and map(q_i) is the best mapping function that permutes clustering labels to match the ground truth labels.

To comprehensively evaluate all methods, we repeated each experiment ten times and reported the median and average clustering errors. In addition, we assessed the robustness of the proposed method by adding noise to each dataset. The noise level is quantified using the Peak Signal-to-Noise Ratio (PSNR) [37], defined as follows:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{s^2}{(1/mn)\sum_i^m\sum_j^n(\mathbf{I}_{ij}-\mathbf{K}_{ij})^2}\right), \tag{4.2}$$

where $\mathbf{I}$ is the noise free data, $\mathbf{K} = \mathbf{I} + \mathbf{N}$ is a noisy data and s is the maximum possible value of an element of $\mathbf{I}$. Note that decreasing values of PSNR means increasing amounts of noise.

4.1. Parameter selections

Before conducting the experiments, we briefly discuss the selection of parameters. In JCLLM, the parameter μ is introduced as a small positive constant to ensure numerical stability and avoid division by zero in the iterative reweighting scheme. Specifically, μ is fixed throughout the optimization and defined as $\mu = \mu_c\|\mathbf{X}\|_2$, where $\|\mathbf{X}\|_2$ denotes the spectral norm of the data matrix and μ_c is a scaling factor. To examine the influence of this numerical parameter, we further conduct a sensitivity analysis with respect to μ_c. As shown in Fig. 2, extremely small or large values of μ_c may lead to degraded

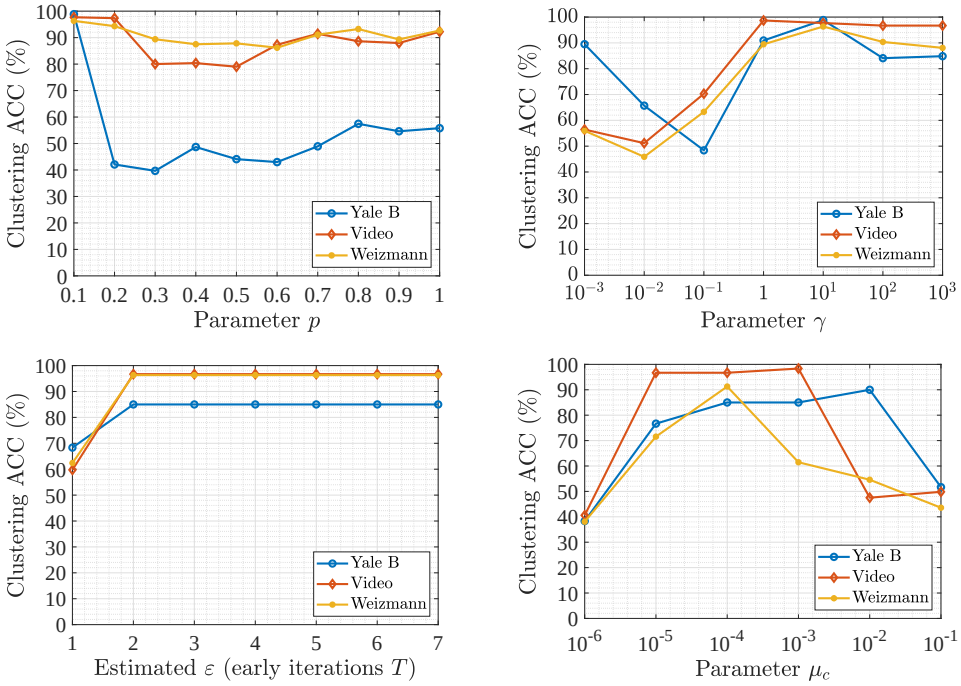


Figure 2: The sensitivity of JCLLM to the model parameters p, γ, ε and μ_c on different databases.

performance, while the clustering accuracy remains stable within a moderate range. Based on this empirical observation, we recommend setting $\mu_c = 10^{-4}$ in practice, which provides a good balance between numerical stability and clustering accuracy.

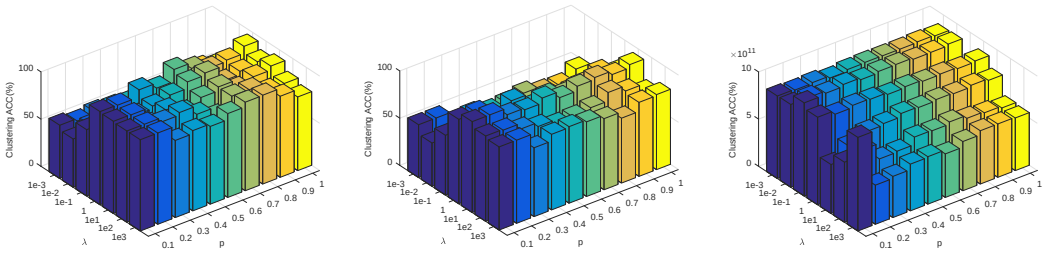
Next, we investigate the sensitivity of JCLLM to the trade-off parameter γ, the power parameter p, and the capped threshold ε. Specifically, γ and p are tuned within the ranges $[10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1, 10, 10^2, 10^3]$ and $[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$, respectively. For the capped threshold, ε is automatically estimated from the residuals in the early iterations and then fixed for the remaining optimization process. To evaluate the robustness of this strategy, we analyze the sensitivity of JCLLM to different early-stage estimation choices, where ε is estimated using varying numbers of initial iterations. Fig. 2 reports the clustering accuracy of JCLLM under different values of p, γ, and ε on multiple datasets. It can be observed that smaller values of p generally lead to more accurate solutions due to the enhanced sparsity-promoting effect of the $\ell_{2,p}$ regularizer. Meanwhile, increasing γ tends to improve clustering accuracy by strengthening temporal consistency. Regarding the capped threshold, the performance stabilizes once ε is estimated from a small number of early iterations, indicating that the proposed strategy is insensitive to the exact early choice of ε.

Fig. 3 shows the accuracy of JCLLM on different datasets. From this, we observe that satisfactory clustering results can be achieved by setting $(\gamma, p) = (10, 0.1)$ for the Video dataset, $(1000, 0.1)$ for the Weizmann dataset, and $(10, 0.1)$ for the Yale B dataset, respectively. In our experiments, the value of ε is initially selected based on the values of the matrix during the first five iterations. Afterward, ε is fixed as a constant. In this way, our model can automatically learn an appropriate value for ε in different situations.

In Algorithm 3.1, we adopt the following stopping criterion:

$$\frac{\|\mathbf{Z}^{(k+1)} - \mathbf{Z}^{(k)}\|_F^2}{\|\mathbf{Z}^{(k+1)}\|_F^2} < \epsilon$$

with $\epsilon = 10^{-4}$.



(a) Video dataset (b) Human activity dataset (c) Yale B dataset

Figure 3: The sensitivity of JCLLM to the model parameters λ and p on different databases.

4.2. JCLLM for subspace segmentation

In this section, we present numerical results of JCLLM and other state-of-the-art algorithms, including SSC [10], LRR [27], SpatSC [16], OSC [37], TSC [25], and TS^{3}C [29]. All experiments are conducted on a PC equipped with an Intel(R) Core(TM) i5-9300H CPU at 2.40 GHz, 8 GB of memory, running Windows 10 and Matlab version R2016b.

1) Synthetic Data Example: In this section, we construct $k = 8$ independent subspaces $\{S_i\}_{i=1}^k$, whose bases $\{\mathbf{U}_i\}_{i=1}^k$ are generated by $\mathbf{U}_{i+1} = \mathbf{T}\mathbf{U}_i$, with $1 \leq i \leq k$, where $\mathbf{T}$ is a random rotation matrix and $\mathbf{U}_i \in \mathbb{R}^{d\times r}$ is a random orthogonal matrix. Thus, each subspace has a rank of $r = 8$, and the data has an ambient dimension of $d = 200$. We sample $n_i = 20$ data vectors from each subspace, where $\mathbf{X}_i = \mathbf{U}_i\mathbf{Q}_i$, and $\mathbf{Q}_i \in \mathbb{R}^{8\times 20}$ is a random Gaussian multivariate matrix. This mimics the assumption that consecutive data points within the same subspace are similar to each other. Finally, the data is concatenated as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k]$.

We repeated the tests 10 times. The clustering accuracies, including both median and mean values, for our approach and all compared methods are shown in Table 2. We also provide visual comparisons of the affinity matrix $\mathbf{Z}$ in Fig. 4 and the clustering results in Fig. 5 for the corresponding affinity matrix. It is worth noting that our method always obtains the true segmentation, regardless of PSNR. Although SSC, SpatSC, and OSC also perform well, our affinity graph exhibits a better block-diagonal structure.

2) Video Segmentation: To evaluate the effectiveness of our JCLLM for video clustering, we conducted experiments on the video dataset [13][†], which contains 10 uncompressed AVI clips of natural scenes with 12 fps, including at least one target objects or something others, Length varies 5-10 seconds. We selected three sequences randomly. Preprocessing of each sequence includes converting the color video to grayscale and sampling to a resolution of 288×352, so every data point can be recorded as $\mathbf{x}_i \in \mathbb{R}^{101376}$. We evaluated the clustering performance of our proposed method and the compared ones.

Every tests were repeated 10 times, and the clustering accuracies with median and mean data were reported in Table 3. For clarity, we visually compared the affinity matrix $\mathbf{Z}$ and the clustering results, as shown in Figs. 6 and 7. Form this, we can draw the following conclusions: Our JCLLM method significantly outperformed all the other compared methods on all the three subjects.

3) Action Clustering: The aim of this experiment is to segment individual scenes from the Weizmann dataset[‡]. The Weizmann dataset contains 90 video sequences (180×144 pixels, 50 fps) captured from 9 subjects [15]. Each subject performs 10 different actions, including jumping-jack (or shortly jack), jump-forward-two-legs (jump),

[†] http://www.kecl.ntt.co.jp/people/kimura.akisato/saliency3.html
[‡] http://www.wisdom.weizmann.ac.il/∼vision/SpaceTimeActions.html

Table 2: Subspace clustering accuracy (%) for the synthetic data set with various magnitudes of Gaussian noise, higher is better.

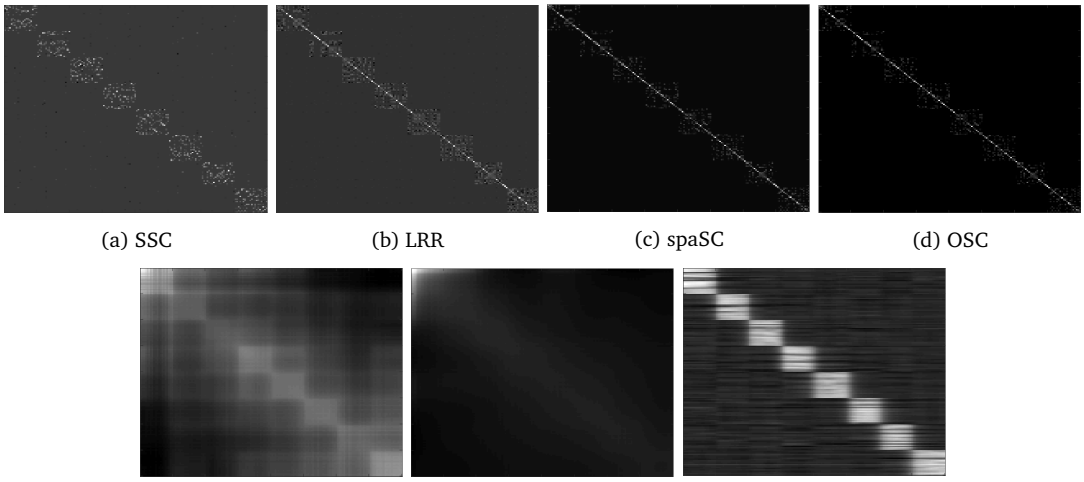
PSNR	ACC	SSC	LRR	spatSC	OSC	TSC	TS^{3}C	JCLLM
Max	Med	**100**	84.69	**100**	**100**	**100**	81.88	**100**
	Mean	**100**	86.37	**100**	**100**	99.69	81.69	**100**
91	Med	**100**	86.25	**100**	**100**	99.68	79.38	**100**
	Mean	**100**	85.5	**100**	**100**	99.56	80.13	**100**
72	Med	**100**	86.87	**100**	**100**	**100**	82.50	**100**
	Mean	**100**	86.87	**100**	**100**	99.93	82.38	**100**
51	Med	**100**	84.37	**100**	**100**	**100**	81.88	**100**
	Mean	**100**	85.68	**100**	**100**	99.75	81.94	**100**
31	Med	**100**	86.25	**100**	**100**	99.37	82.50	**100**
	Mean	**100**	86.06	**100**	**100**	99.62	82.19	**100**

Table 3: Subspace clustering accuracy (%) for the video segmentation dataset [13] with various magnitudes of Gaussian noise, higher is better. Experiments show that JCLLM achieve better segmentation.

PSNR	ACC	SSC	LRR	spatSC	OSC	TSC	TS^{3}C	JCLLM
Max	Med	55.76	42.57	61.91	64.88	77.95	93.07	**97.03**
	Mean	57.06	43.43	61.97	64.69	78.02	93.73	**96.96**
81	Med	55.76	43.56	61.91	64.88	77.95	91.09	**96.68**
	Mean	57.06	44.15	61.89	63.96	74.95	84.42	**95.05**
41	Med	55.76	45.21	61.91	61.25	78.28	86.14	**98.35**
	Mean	57.06	44.98	61.91	63.04	78.48	88.84	**98.09**
27	Med	55.76	89.10	66.14	60.92	77.95	86.14	**97.03**
	Mean	57.06	89.63	66.20	61.65	78.28	85.87	**97.43**

jump-in-place-on-two-legs (pjump), gallop-sideways (side), bend, skip, walk, run, wave-one hand (wave1), and wave-two-hands (wave2). Fig. 8 shows 10 frames of different actions in the dataset.

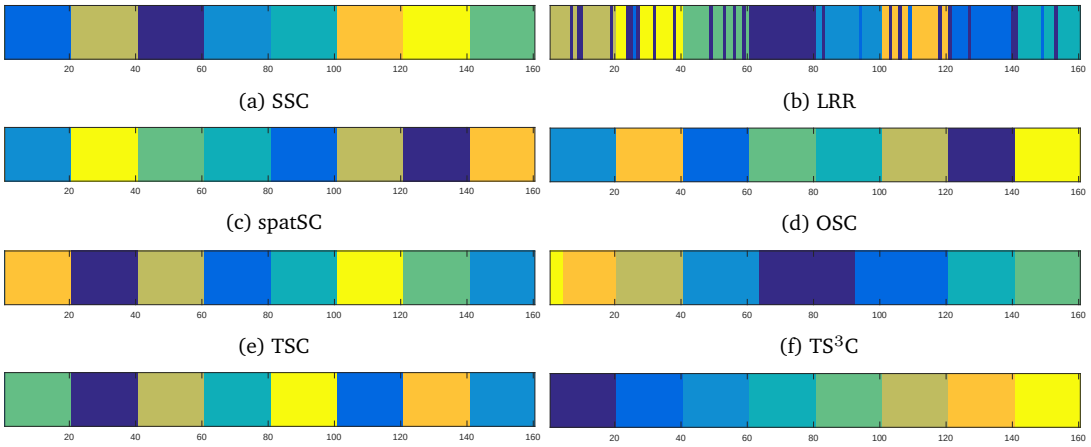
The following experiments were conducted on action clips from 3 randomly chosen subjects. Similar to the previous experiments, we add increasing Gaussian noise and repeat the experiments 10 times for each noise magnitude. The statistical results of segmenting 3 action sequences are reported in Table 4, and the corresponding visual results are presented in Figs. 9 and 10. Specifically, it can be observed that JCLLM is the most accurate method. We observe that SSC, and especially LRR, fail to obtain meaningful temporal segments, as they do not take into account the temporal information. This demonstrates that properly designed local temporal regularization $f(\mathbf{Z})$ is more important than global structural regularization $\delta(\mathbf{Z})$.



(a) SSC (b) LRR (c) spaSC (d) OSC

(e) TSC (f) TS^{3}C (g) JCLIM

Figure 4: The affinity matrix of the synthesized data is obtained without adding noise.



(a) SSC (b) LRR

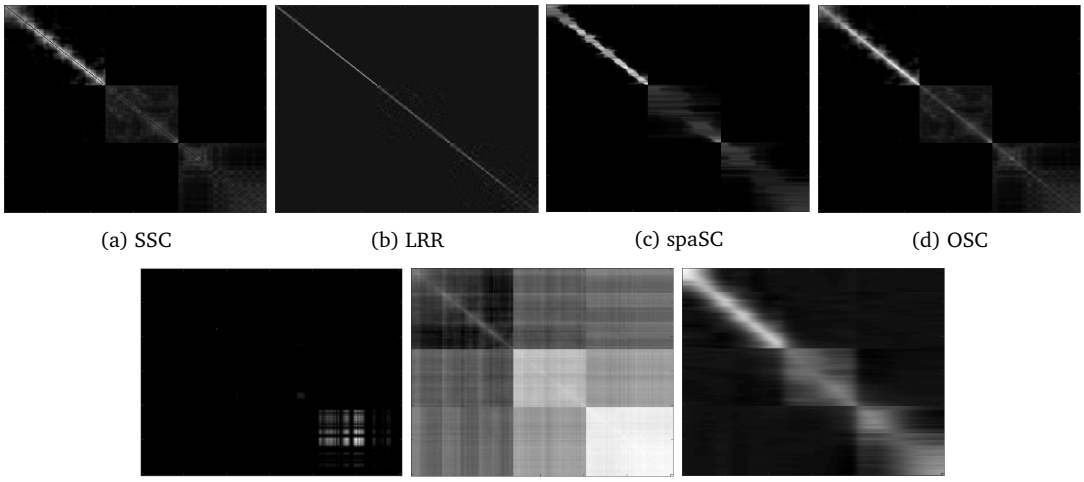
(c) spatSC (d) OSC

(e) TSC (f) TS^{3}C

(g) JCLLM (h) Ground Truth

Figure 5: Clustering results from affinity matrices in Fig. 4. JCLLM, SSC, OSC and spatSC achieve perfect segmentation, while LRR and TSC suffer from misclassification.

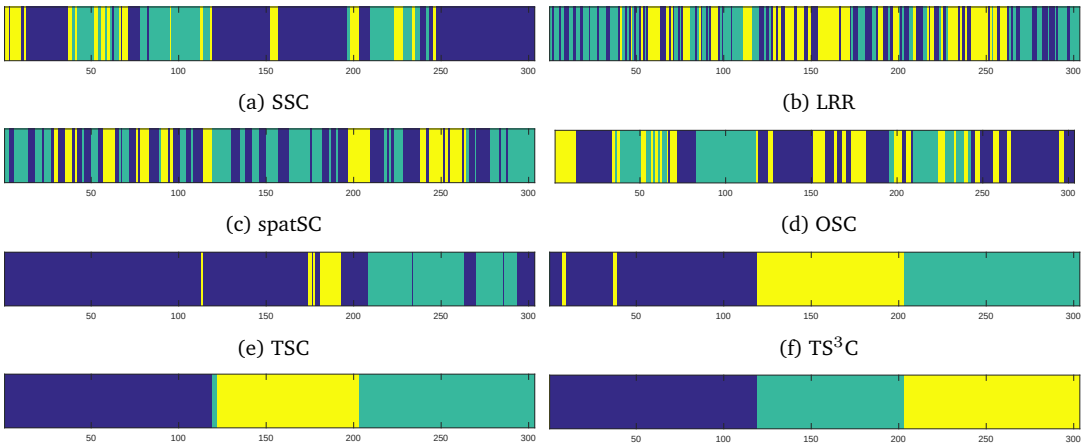
4) Face Clustering: In this subsection, following the setup of SSC-L1TG [18], we conducted experiments on the Extended Yale Face Database B [14], with the goal of segmenting or clustering unique subjects from a set of face images. Although this task is not exactly suited to our assumption of clustering sequential data, we ensure that in our datasets, face images are kept contiguous, i.e., unique subjects do not mix with each other. We can exploit the temporal information, as the neighbors of each data point are likely to belong to the same subject. As shown in Fig. 11, three subjects are taken as examples. For each test, we randomly select three subjects from the dataset, and then 20 sample images are randomly chosen from each subject. To highlight the



(a) SSC (b) LRR (c) spaSC (d) OSC

(e) TSC (f) TS^{3}C (g) JCLIM

Figure 6: The affinity matrix of the video data is obtained without adding noise.



(a) SSC (b) LRR

(c) spatSC (d) OSC

(e) TSC (f) TS^{3}C

(g) JCLLM (h) Ground Truth

Figure 7: Clustering results from affinity matrices in Fig. 6. JCLLM achieves better segmentation, while SSC,LRR, spatSC, TSC and OSC suffer from misclassification.

efficiency of our method, the images are directly used as data vectors $\mathbf{x}_i \in \mathbb{R}^{307200}$ without dimensionality reduction. We then concatenate them to ensure that subjects are not mixed. We repeated these tests 10 times for each level of corruption by Gaussian noise.

Table 5 presents the subspace clustering results on the Extended Yale B database. From the results, we observe that our method outperforms state-of-the-art methods in these subspace clustering tasks. Specifically, for a PSNR of 85, our method surpasses the second-best result by 16.1%. Compared to other methods that use the Frobenius norm to handle noise, our method achieves better performance. Therefore, our proposed method is more robust to noise and more effective for subspace clustering.

Table 4: Subspace clustering accuracy (%) for the 3 motion sequences with various magnitudes of Gaussian noise, higher is better, experiments show that JCLLM achieve better segmentation.

PSNR	ACC	SSC	LRR	spatSC	OSC	TSC	TS^{3}C	JCLLM
Max	Med	65.60	59.17	87.16	66.06	73.70	78.90	**96.33**
	Mean	68.07	59.36	87.52	68.44	73.61	78.90	**95.49**
40	Med	66.97	60.09	88.07	66.06	77.16	81.19	**96.33**
	Mean	72.39	66.88	88.99	72.11	76.92	80.73	**96.32**
20	Med	65.60	59.17	87.16	66.06	76.70	70.64	**96.33**
	Mean	68.07	59.36	87.52	68.44	76.61	70.37	**95.48**
4	Med	65.60	59.17	87.16	66.06	76.69	38.99	**96.33**
	Mean	68.07	59.36	87.52	68.44	76.61	40.55	**95.48**

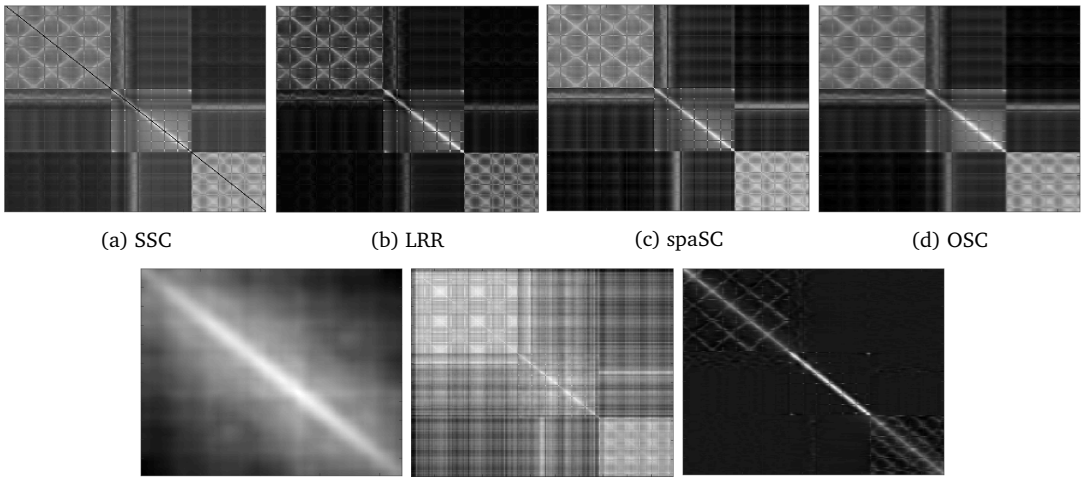
Table 5: Subspace clustering accuracy (%) for face data with various magnitudes of Gaussian noise, higher is better, experiments show that JCLLM achieve better segmentation.

PSNR	ACC	SSC	LRR	spatSC	OSC	TSC	TS^{3}C	JCLLM
Max	Med	66.67	70.00	68.33	75.00	75.00	55.00	**85.00**
	Mean	67.67	70.33	68.33	71.00	75.67	54.00	**85.91**
85	Med	68.33	70.00	68.33	65.00	70.00	53.33	**85.02**
	Mean	68.33	71.00	68.00	65.00	70.33	55.00	**86.43**
45	Med	68.33	71.67	68.33	65.00	71.67	55.00	**85.00**
	Mean	68.33	72.00	68.33	67.00	75.33	56.00	**86.19**
30	Med	66.67	73.33	68.33	66.68	73.33	56.67	**85.31**
	Mean	67.67	72.67	68.33	66.67	74.00	57.00	**84.58**

Walk Run Jump Side Bend

Wave1 Wave2 Pjump Jack Skip

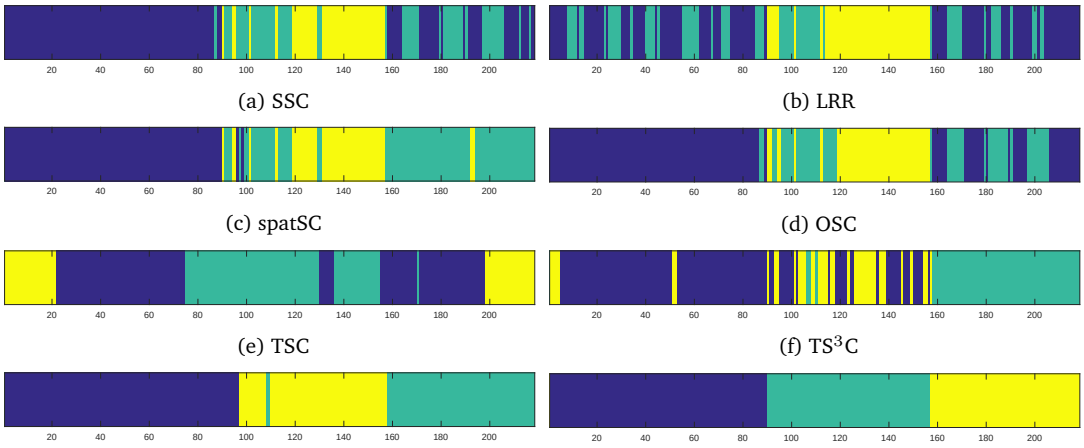
Figure 8: Example frames of the 10 different actions in Weizmann dataset.



(a) SSC (b) LRR (c) spaSC (d) OSC

(e) TSC (f) TS^{3}C (g) JCLIM

Figure 9: The affinity matrix of the three classes human activity data is obtained without adding noise.



(a) SSC (b) LRR

(c) spatSC (d) OSC

(e) TSC (f) TS^{3}C

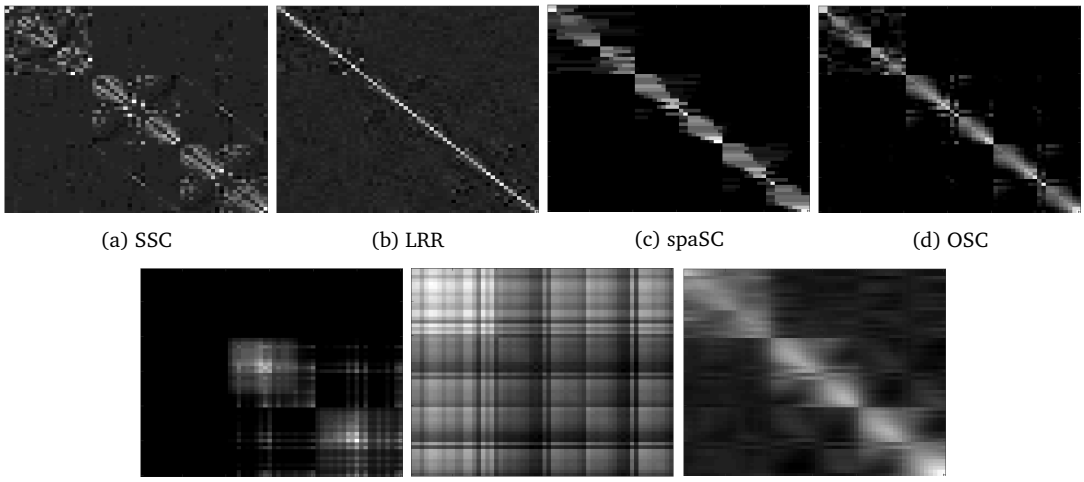
(g) JCLLM (h) Ground Truth

Figure 10: Clustering results from affinity matrices in Fig. 9. JCLLM achieve better segmentation.

Figure 11: Face clustering: given an sequential set of face images (those temporal neighbors are almost all from the same subject), the goal is to cluster images that belong to the same subject.

4.3. Convergence study and computational time

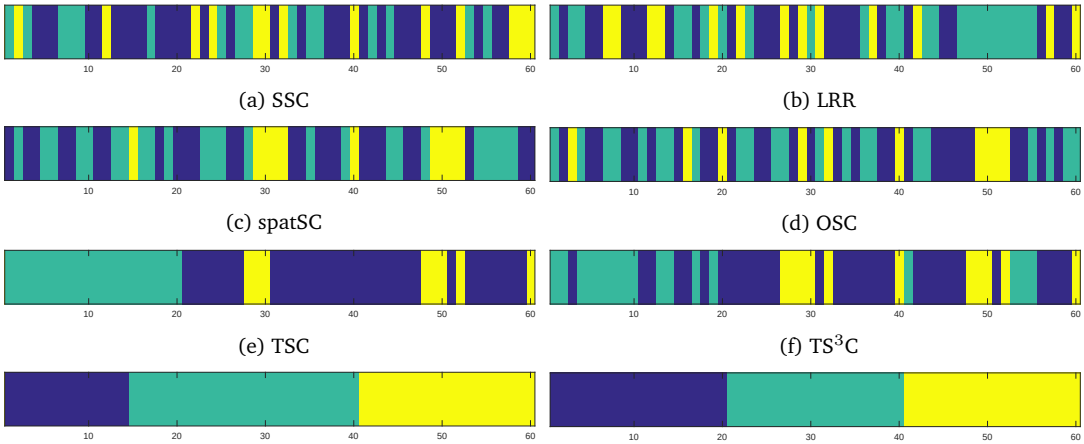
To demonstrate the convergence performance of JCLLM in Algorithm 3.1, we plot the curves of our objective function, shown in Eq. (3.2), with respect to the number of



(a) SSC (b) LRR (c) spaSC (d) OSC

(e) TSC (f) TS^{3}C (g) JCLIM

Figure 12: The affinity matrix of the face data is obtained without adding noise.



(a) SSC (b) LRR

(c) spatSC (d) OSC

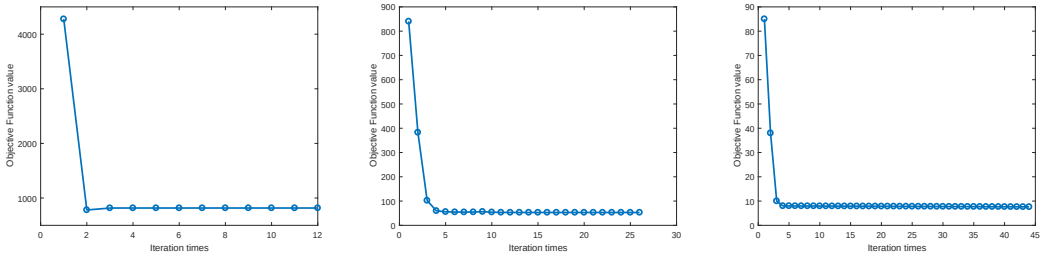
(e) TSC (f) TS^{3}C

(g) JCLLM (h) Ground Truth

Figure 13: Clustering results from affinity matrices in Fig. 12. JCLLM achieve perfect segmentation.

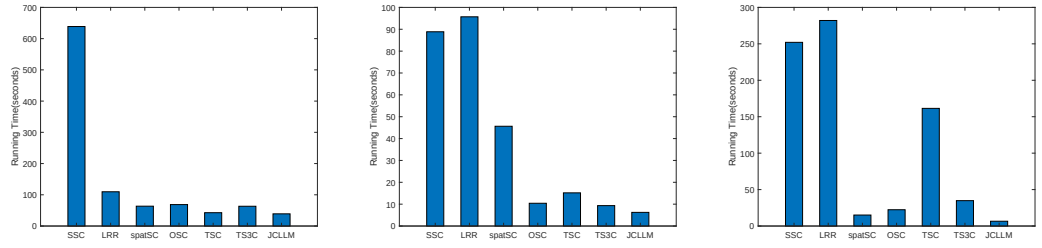
iterations in Fig. 14. As shown in the figure, the optimization algorithm proposed in this paper (Algorithm 3.1) is effective and exhibits fast convergence, with the number of iterations typically less than 45.

In addition, we compare the computation time (in seconds) of the different methods. Since all methods use Ncut for segmentation after obtaining the affinity matrix, we compare only the time taken to compute the representation of the matrix $\mathbf{Z}$ to ensure fairness. The comparison results are shown in Fig. 15. As shown, our method is the fastest across all datasets. This is expected, as our method does not require singular value decomposition, and the system's solution is accelerated by the use of Lyapunov equations. In contrast, SSC, LRR, and TSC are much slower. SSC converges slowly, LRR



(a) Video dataset (b) Weizmann dataset (c) Yale B dataset

Figure 14: Convergence behavior of Algorithm 3.1 on three data sets.



(a) Video dataset (b) Weizmann dataset (c) Yale B dataset

Figure 15: Comparison of computational time (in seconds).

involves SVD, and TSC requires learning a dictionary in each iteration. The complexity of TSC is $\mathcal{O}(nr^3)$ per iteration, which is potentially lower than that of JCLLM, but TSC requires many more iterations, leading to slower performance overall. Thus, our method is more efficient than the others.

5. Conclusions

In this paper, we propose a new robust and efficient sequential subspace clustering method, referred to as JCLLM, which emphasizes the simultaneous minimization of the capped ℓ_2 norm loss and the $\ell_{2,p}$ norm regularizer. The capped ℓ_2 norm-based regression loss function is robust to outliers in datasets, and the $\ell_{2,p}$ norm regularization is used to enforce the temporal properties of sequential data when $0 < p \leq 1$. An efficient approach is employed to solve the problem, and convergence analysis demonstrates the method's convergence behavior. We provide an example using the capped ℓ_2 norm to illustrate its robustness. The proposed JCLLM is tested on various types of data, including synthetic, video, human activity, and face data. The experimental results show that our method outperforms state-of-the-art methods, such as SSC, LRR, SpatSC, OSC, and TSC. In the future, we plan to explore the addition of the capped $\ell_{2,p}$ norm to further enforce the temporal properties. Additionally, we aim to investigate the application of the capped norm to other tasks.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 12501587, 62266002, U21A20455, 12326619), by the Natural Science Foundation of Jiangxi Province of China (Grant Nos. 20252BAC200149, 20252BAC 250007), by the Natural Science Foundation of Guangdong Province of China (Grant Nos. 2023A1515011691, 2024A1515011913), by the Program for Early-career Science and Technology Talents of Jiangxi Province of China (Grant No. 20252BEJ73034), by the Doctoral Science Foundation of Jiangxi Science and Technology Normal University (Grant No. 2024BSQD33), by the National Social Science Foundation of China (Grant No. 22BTJ033), and by the Jiangxi Provincial Higher Education Social Sciences Research Project (Grant No. JD21075).

References

[1] P. BENNER, C. R. LI, AND N. TRUHAR, *On the ADI method for Sylvester equations*, J. Comput. Appl. Math. 233(4) (2009), 1035–1045.

[2] S. BOYD ET AL., *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn. 3(1) (2011), 1–122.

[3] P. S. BRADLEY AND O. L. MANGASARIAN, *k-plane clustering*, J. Glob. Optim. 16(1) (2010), 23–32.

[4] R. CHARTRAND, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett. 14(10) (2007), 707–710.

[5] R. CHARTRAND AND W. YIN, *Iteratively reweighed algorithms for compressive sensing*, in: Proceedings of the 33rd International Conference on Acoustics, Speech, and Signal Processing, (2008), 3869–3872.

[6] H. J. CHE, C. L. LI, M. F. LEUNG, D. OUYANG, X. G. DAI, AND S. P. WEN, *Robust hypergraph regularized deep non-negative matrix factorization for multi-view clustering*, IEEE Trans. Emerg. Top. Comput. Intell. 9(2) (2024), 1817–1829.

[7] L. CHEN, Z. TANG, AND X. FAN, *Enhancing video segmentation with contrastive self-supervised learning of distinctive class features for visually homogeneous frames*, Expert Syst. Appl. 271 (2025), Paper No. 126594.

[8] J. P. COSTEIRA AND T. KANADE, *A multibody factorization method for independently moving objects*, Int. J. Comput. Vis. 29(3) (1998), 159–179.

[9] M. DING, X. L. ZHAO, J. H. YANG, Z. ZHOU, AND M. K. NG, *Bilateral tensor low-rank representation for insufficient observed samples in multidimensional image clustering and recovery*, SIAM J. Imaging Sci. 18(1) (2025), 20–59.

[10] E. ELHAMIFAR AND R. VIDAL, *Sparse subspace clustering*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2009), 2790–2797.

[11] E. ELHAMIFAR AND R. VIDAL, *Sparse subspace clustering: Algorithm, theory, and applications*, IEEE Trans. Pattern Anal. Mach. Intell. 35(11) (2013), 2765–2781.

[12] C. EMMANUEL, M. B. WAKIN, AND S. P. BOYD, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl. 14(5) (2007), 877–905.

[13] K. FUKUCHI, K. MIYAZATO, A. KIMURA, S. TAKAGI, AND J. YAMATO, *Saliency-based video segmentation with graph cuts and sequentially updated priors*, in: Proceedings of the IEEE International Conference on Multimedia and Expo, (2009), 638–641.

[14] A. GEORGHIADES, P. BELHUMEUR, AND D. KRIEGMAN, *From few to many: Illumination cone models for face recognition under variable lighting and pose*, IEEE Trans. Pattern Anal. Mach. Intell. 23(6) (2001), 643–660.

[15] L. GORELICK, M. BLANK, E. SHECHTMAN, M. IRANI, AND R. BASRI, *Actions as space-time shapes*, IEEE Trans. Pattern Anal. Mach. Intell. 29(12) (2007), 2247–2253.

[16] Y. GUO, J. B. GAO, AND F. LI, *Spatial subspace clustering for hyperspectral data segmentation*, in: Proceedings of the Conference of the Society of Digital Information and Wireless Communications, (2013), 180–190.

[17] T. HASTIE AND P. SIMARD, *Metrics and models for handwritten character recognition*, Stat. Sci. 13(1) (1998), 54–65.

[18] W. HU, S. LI, W. ZHENG, Y. LU, AND G. H. YU, *Robust sequential subspace clustering via ℓ_1-norm temporal graph*, Neurocomputing 383 (2020), 380–395.

[19] W. HU, Y. LU, AND J. REN, *A fixed-point proximity algorithm for recovering low-rank components from incomplete observation data with application to motion capture data refinement*, J. Comput. Appl. Math. 410 (2022), Paper No. 114224.

[20] W. HU, Z. WANG, S. LIU, X. YANG, G. YU, AND J. J. ZHANG, *Motion capture data completion via truncated nuclear norm regularization*, IEEE Signal Process. Lett. 25(2) (2017), 258–262.

[21] W. HU, X. ZHU, T. WANG, Y. YI, AND G. YU, *Discrete subspace structure constrained human motion capture data recovery*, Appl. Soft Comput. J. 129 (2022), Paper No. 109617.

[22] Z. JIANG, Z. LIN, AND L. DAVIS, *Recognizing human actions by learning and matching shape-motion prototype trees*, IEEE Trans. Pattern Anal. Mach. Intell. 34(3) (2012), 533–547.

[23] G. LAN, C. HOU, AND D. YI, *Robust feature selection via simultaneous capped ℓ_2-norm and $\ell_{2,1}$-norm minimization*, in: Proceedings of the IEEE International Conference on Big Data Analysis, 2016.

[24] B. LI, Y. ZHANG, Z. LIN, AND H. LU, *Subspace clustering by mixture of Gaussian regression*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2015), 2094–2102.

[25] S. LI, K. LI, AND Y. FU, *Temporal subspace clustering for human motion segmentation*, in: Proceedings of the IEEE International Conference on Computer Vision, (2015), 4453–4461.

[26] J. LIN, X. GAO, Z. ZHANG, AND H. DENG, *DASC: Learning discriminative latent space for video clustering*, Neurocomputing 637 (2025), Paper No. 130050.

[27] G. C. LIU, Z. C. LIN, AND Y. YU, *Robust subspace segmentation by low-rank representation*, in: Proceedings of the 27th International Conference on Machine Learning, (2010), 663–670.

[28] G. LIU AND S. YAN, *Latent low-rank representation for subspace segmentation and feature extraction*, in: Proceedings of the 2011 International Conference on Computer Vision, (2011), 1615–1622.

[29] H. J. LIU, J. CHENG, AND F. WANG, *Sequential subspace clustering via temporal smoothness for sequential data segmentation*, IEEE Trans. Image Process. 27(2) (2018), 866–878.

[30] C. Y. LU ET AL., *Robust and efficient subspace segmentation via least square regression*, in: Proceedings of the European Conference on Computer Vision, (2012), 347–360.

[31] Y. LU ET AL., *Projective robust nonnegative factorization*, Inf. Sci. 364 (2016), 16–32.

[32] Y. MA, H. DERKSEN, W. HONG, AND J. WRIGHT, *Segmentation of multivariate mixed data via lossy data coding and compression*, IEEE Trans. Pattern Anal. Mach. Intell. 29(9) (2007), 1546–1562.

[33] F. NIE, Z. HUO, AND H. HUANG, *Joint capped norms minimization for robust matrix recovery*, in: Proceedings of the International Joint Conference on Artificial Intelligence, (2017), 2557–2563.
[34] X. Y. PU, H. J. CHE, B. C. PAN, M. F. LEUNG, AND S. P. WEN, *Robust weighted low-rank tensor approximation for multiview clustering with mixed noise*, IEEE Trans. Comput. Soc. Syst. 11(3) (2023), 3268–3285.
[35] J. B. SHI AND J. MALIK, *Normalized cuts and image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell. 22(8) (2000), 888–905.
[36] J. V. STONE, *Blind source separation using temporal predictability*, Neural Comput. 13(7) (2001), 1559–1574.
[37] S. TIERNEY, J. B. GAO, AND Y. GUO, *Subspace clustering for sequential data*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2014), 1019–1026.
[38] P. TSENG, *Convergence of a block coordinate descent method for nondifferentiable minimization*, J. Optim. Theory Appl. 109(3) (2001), 475–494.
[39] R. VIDAL, Y. MA, AND S. SASTRY, *Generalized principal component analysis*, IEEE Trans. Pattern Anal. Mach. Intell. 27(12) (2005), 1945–1959.
[40] T. ZHANG, X. TIAN, Y. ZHOU, S. JI, X. WANG, X. TAO, AND Y. WU, *Dvis++: Improved decoupled framework for universal video segmentation*, IEEE Trans. Pattern Anal. Mach. Intell. 47(7) (2025), 5918–5929.
[41] X. J. ZHANG, C. XU, X. L. SUN, AND G. BACIU, *Schatten-q regularizer constrained low rank subspace clustering model*, Neurocomputing 182 (2016), 36–47.
[42] P. ZHOU, C. LU, J. FENG, Z. LIN, AND S. YAN, *Tensor low-rank representation for data recovery and clustering*, IEEE Trans. Pattern Anal. Mach. Intell. 43(5) (2019), 1718–1732.