Automatic Model Selection in Subspace Clustering via Triplet Relationships

Jufeng Yang[1], Jie Liang[1], Kai Wang[1], Yong-Liang Yang[2], Ming-Ming Cheng[1]
[1]College of Computer and Control Engineering
Nankai University
No.38 Tongyan Road, Tianjin, China
[2]Department of Computer Science
University of Bath
Claverton Down, Bath, United Kingdom

Abstract

This paper addresses both the model selection (*i.e.* estimating the number of clusters K) and subspace clustering problems in a unified model. The real data always distribute on a union of low-dimensional sub-manifolds which are embedded in a high-dimensional ambient space. In this regard, the state-of-the-art subspace clustering approaches firstly learn the affinity among samples, followed by a spectral clustering to generate the segmentation. However, arguably, the intrinsic geometrical structures among samples are rarely considered in the optimization process. In this paper, we propose to simultaneously estimate K and segment the samples according to the local similarity relationships derived from the affinity matrix. Given the correlations among samples, we define a novel data structure termed the Triplet, each of which reflects a high relevance and locality among three samples which are aimed to be segmented into the same subspace. While the traditional pairwise distance can be close between inter-cluster samples lying on the intersection of two subspaces, the wrong assignments can be avoided by the hyper-correlation derived from the proposed triplets due to the complementarity of multiple constraints. Sequentially, we propose to greedily optimize a new model selection reward to estimate K according to the correlations between inter-cluster triplets. We simultaneously optimize a fusion reward based on the similarities between triplets and clusters to generate the final segmentation. Extensive experiments on the benchmark datasets demonstrate the effectiveness and robustness of the proposed approach.

Introduction

State-of-the-art subspace clustering methods model high-dimensional data samples $\boldsymbol{X} = \{\boldsymbol{x}_i \in \mathbb{R}^D\}_{i=1}^N$ into a union of low-dimensional subspaces $\{S_j\}_{j=1}^K$ (Elhamifar and Vidal 2009; Vidal 2011), where D and N denote the dimensionality and scale of the given dataset, respectively, $K \geq 1$. The primary step of the spectral based subspace clustering methods is to calculate the coefficient matrix $\boldsymbol{C}$ by solving an optimization problem as follows:

$$\min_{\boldsymbol{C}} \ L(\boldsymbol{X}\boldsymbol{C}, \boldsymbol{X}) + \lambda \|\boldsymbol{C}\|_\xi, \tag{1}$$

where $L(\cdot,\cdot) : \mathbb{R}^{N\times N} \rightarrow \mathbb{R}^+$ denotes the reconstruction loss, λ is the trade-off parameter and $\|\cdot\|_\xi$ denotes the regularization term where different ξ's lead to ℓ_0, ℓ_1, ℓ_2, ℓ_∞ or the nuclear norm (Vidal 2011; Yang et al. 2016). The algorithm then employs the spectral clustering (Shi and Malik 2000) on the affinity matrix derived from $\boldsymbol{C}$ for final assignments. Note that in practice, both the number of subspaces K and their dimensions $\{d_j\}_{j=1}^K$ are always unknown (Vidal 2011; Wang and Zhu 2015). Hence, the goals of subspace clustering include finding the appropriate K and assigning data points into K clusters (Elhamifar and Vidal 2013; Li et al. 2016).

However, modeling the cluster number in the optimization framework is difficult (Li et al. 2016), since the definition of clusters is unquantifiable on the complex data space. Sequentially, most of the spectral based subspace clustering algorithms set the parameter K manually, which achieve state-of-the-art performance on the confined applications where the number of clusters is fixed and given.

Clustering aims to group the similar patterns into the same cluster by maximizing the inter-cluster dissimilarity and the intra-cluster similarity. An effective way for estimating K is to map the original samples into intrinsic correlation space, followed by an iterative optimization according to the local similarity relationships among samples. Elhamifar *et al.* (Elhamifar and Vidal 2013) propose that the $\boldsymbol{\Gamma}\boldsymbol{C}$ achieves a block-diagonal structure with K blocks corresponding to K clusters, where $\boldsymbol{C}$ is derived from (1) and $\boldsymbol{\Gamma}$ denotes a proper permutation matrix. Moreover, Peng *et al.* (Peng et al. 2016) verify the intra-subspace projection dominance (IPD) of $\boldsymbol{C}$ derived from (1) with various ξ's, *i.e.*, for all $\boldsymbol{x}_p, \boldsymbol{x}_q \in S$ and $\boldsymbol{x}_k \notin S$, we have $c_{pq} \geq c_{pk}$. Hence, considering a graph $\boldsymbol{G} = (\boldsymbol{X}, \boldsymbol{C})$ (Nasihatkon and Hartley 2011), *i.e.*, $\boldsymbol{x}_i, \boldsymbol{x}_j \in \boldsymbol{X}$ being the vertices and c_{ij} being the weight between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the automatic segmentation can be greedily generated via the following two steps inspired by the density based algorithms (Rodriguez and Laio 2014): 1) finding $\widehat{K}$ initialized clusters $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$ by minimizing the weighted sum of all inter-cluster connections. 2) assigning the remaining samples $\boldsymbol{x}$ to a proper $\mathcal{C}$ by maximizing the weighted connections between $\boldsymbol{x}$ and $\mathcal{C}$.

Yet, since the low-dimensional sub-manifolds can be very dense (Peng, Zhang, and Yi 2013), the points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ which are close regarding the pairwise distance may not belong to the same subspace, especially near the intersection of two subspaces. A hypergraph $\widehat{\boldsymbol{G}}$ in which one edge can link up more than two vertices (Gao, Tsang, and Chia 2013;

Kim et al. 2014) is then proposed to replace the pairwise $\boldsymbol{G}$. Besides, the local geometry of each $\boldsymbol{x}_i$ can be linearly reconstruct from its correlated points (*i.e.* $\boldsymbol{x}_j$ for which the c_{ij} is large) (Yin, Gao, and Lin 2016). In this paper, we further introduce a novel data structure termed the *Triplet*, *i.e.*, $\boldsymbol{\tau}$, to explore the local geometry with hyper-correlations rather than pairwise similarity. Each $\boldsymbol{\tau}$ contains three points, *i.e.*, $\boldsymbol{x}_i$, $\boldsymbol{x}_j$, $\boldsymbol{x}_k$, together with their correlations, *i.e.*, c_{ij}, c_{jk} and c_{ki}, for which we considered as a meta-element for clustering. All the correlations are large enough to ensure that the three points are strongly connected according to the IPD property (Peng et al. 2016). The proposed triplet relationship achieves favorable theoretical guarantee against the pairwise one in the following two folds. On one hand, it is more robust when partitioning the inter-cluster samples near the intersection of two sub-manifolds due to the complementarity of multiple constraints. On the other hand, it evokes mutual restrictions of neighbored samples thus depicts a local geometrical structure, by which we can calculate the segmentation greedily.

Sequentially, in this paper, we propose a framework termed the autoSC to simultaneously estimate the number of clusters and segment the samples by exploring the local geometrical structure derived from the given coefficient matrix $\boldsymbol{C}$. Specifically, we first generate $\boldsymbol{C}$ by solving the subspace representation problem in (1), followed by extracting the triplet relationship $\boldsymbol{\tau}$. Then, we greedily initialize $\widehat{K}$ clusters, *i.e.*, $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$, to maximize the inter-cluster dissimilarity among triplets, which is obtained via a new model selection reward. Finally, we assign each of the remaining samples into $\mathcal{C}$ to maximize the intra-cluster similarity by optimizing a new fusion reward.

We have mainly two contributions. First, we define the triplet relationship $\boldsymbol{\tau}$ which induces a high relevance and locality among three samples to explore the local similarity, and verify its favorable performance against the traditional pairwise relation. Second, we design a greedy framework for joint model selection and clustering utilizing the intrinsic geometrical structures depicted by the proposed triplet relationships. Extensive experiments on benchmarks indicate that our proposed autoSC outperforms the state-of-the-art methods.

Related Work

Subspace Representations

Automatically finding the clusters of samples is a crucial issue in computer vision (Elhamifar and Vidal 2009; Zhao et al. 2017; Schroff, Kalenichenko, and Philbin 2015). Traditional subspace clustering first constructs a sparse linear representation of each data sample using the remaining as a dictionary (Li and Vidal 2015; Cheng et al. 2016). It generates a coefficient matrix $\boldsymbol{C} \in \mathbb{R}^{N\times N}$ by solving the optimization problem in (1). By modifying the regularization term, *i.e.*, the value of ξ in $\|\boldsymbol{C}\|_\xi$, researchers generate coefficient matrix $\boldsymbol{C}$ considering different intrinsic properties of the data space $\boldsymbol{X}$.

The ℓ_0 and ℓ_1 norm based algorithms (Yang et al. 2016) intend to eliminate most non-zero values of $\boldsymbol{C}$, since c_{ij} denotes the similarity between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and the connections only exist when $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ belong to same subspace. Meanwhile, there are ℓ_2 and nuclear norm based methods (Hu et al. 2014; Liu, Latecki, and Yan 2010) intending to preserve as many the non-zero values in $\boldsymbol{C}$. These frameworks interpret the coefficient matrix by that each column of $\boldsymbol{C}$, *i.e.*, $\boldsymbol{c}_i$, is the self-representation of $\boldsymbol{x}_i$, thus they preserve all values to ensure the mapping invariance (also termed as the grouping effect (Lu et al. 2012)). Recently, the mixed norms such as trace Lasso (Lu et al. 2015) and elastic net (You et al. 2016) are applied to the optimization model for the tradeoff between the sparsity and the grouping effect. There also arise frameworks which incorporate various constraints into the model. For example, (Guo, Gao, and Li 2013) considers the similarity between neighbors in sequential data, of which the new penalty $\|\boldsymbol{CR}\|_1$ forces consecutive columns of $\boldsymbol{C}$ to be similar where $\boldsymbol{R}$ is a lower triangular matrix with -1 on the diagonal and 1 on the second diagonal.

In this paper, we generate the triplet relationship by exploring the intrinsic geometrical structures depicted in the coefficient matrix $\boldsymbol{C}$ from off-the-shelf subspace representation modules. We verify that the proposed method is robust to the combination of $\boldsymbol{C}$ with various properties.

Estimating the Number of Clusters

However, in most real applications, K is unknown to the clustering algorithms (Elhamifar and Vidal 2009). Considering the block-diagonal structure of the coefficient matrix (Feng et al. 2014), Liu *et al.* (Liu et al. 2013) propose to utilize the heuristic estimator: $\hat{K} = N - \text{round}(\sum f_\epsilon(\sigma_i))$, where ϵ is a cut-off threshold, σ_i denotes a singular value of normalized Laplacian matrix and f_ϵ is a summation function which counts different values regarding that $\sigma_i < \epsilon$. The singular based methods (Favaro, Vidal, and Ravichandran 2011; Elhamifar and Vidal 2009) rely on the large gap between singulars, which is effective only when the sub-manifolds are sparse in the ambient space.

The density based methods always iteratively find the optimal number of clusters and the optimal assignment by analyzing the density among samples. In (Rodriguez and Laio 2014), Rodriguez *et al.* assume that the cluster centers are characterized by a higher density than its neighbors and different centers should be far enough. For each sample $\boldsymbol{x}_i$, its local density $\rho_i = \sum_j \mathcal{X}(d_{ij} - d_c)$ and the distance $\delta_i = \min_{\boldsymbol{x}_j:\rho_j>\rho_i}(d_{ij})$ from points of higher density are iteratively calculated for comparison. The algorithm finds a tradeoff between ρ and δ to decide the cluster centers and the assignment of the remained samples. Wang *et al.* (Wang and Zhu 2015) propose the Dirichlet process based Bayesian nonparametric method (DP-space) to exploit the tradeoff between data fitness and model complexity, which is more tolerate to noisy and outlier values than the alternative algebraic and geometry solutions. Li *et al.* propose the SCAMS (Li, Cheong, and Zhou 2014) which penalizes the clustering cost by minimizing the Frobenius inner product $-\langle \boldsymbol{C}, \boldsymbol{\mathcal{B}}\rangle$ and estimates K by minimizing $rank(\boldsymbol{\mathcal{B}})$, where $\boldsymbol{\mathcal{B}}$ is a binary relationship matrix encoding the pairwise relationships among samples. Correlation clustering (CC) (Beier, Hamprecht, and Kappes 2015) minimizes the sum of the weights of the cut edges on

an undirected graph with positive and negative edge weights, such that $y^* = \arg\min_y \sum w_{ij} y_{ij}$ where the label $y_{ij} = 1$ as cut and $y_{ij} = 0$ as uncut. However, these algorithms only take the pairwise correlation into consideration, and they can not be robust when subspaces are very dense and samples of different subspaces are not completely distinguished.

The hyper-graph relation (Lu et al. 2016; Purkait et al. 2014) avoids such drawbacks and the literature follows two different directions. Some transform the hyper relationship into another pairwise graph (Gao, Tsang, and Chia 2013; Schölkopf, Platt, and Hofmann 2006), followed by the conventional graph clustering method (Shi and Malik 2000) to generate the segmentation. There are also generalization methods (Liu, Latecki, and Yan 2010; Li et al. 2016) extending the pairwise graph to the hyper-graph or tensor analysis. For example, the tensor affinity variant of SCAMS, *i.e.*, SCAMSTA (Li et al. 2016), exploits the higher order mathematical structures by providing multi groups of nodes in $\boldsymbol{\mathcal{Z}}$, *i.e.*, $\boldsymbol{\mathcal{Z}} = \sum_k \boldsymbol{z}_r \circ \boldsymbol{z}_r \circ \cdots \circ \boldsymbol{z}_r \in \{0,1\}^{N \times N \times \cdots \times N}$, where $\boldsymbol{z}_r \circ \boldsymbol{z}_r$ denotes the outer product and $\boldsymbol{z}_r \in \{0,1\}^N$ is the indicator vector.

In this paper, we estimate K by initializing clusters with optimal inter-cluster dissimilarities. We obtain the estimation by exploring the local correlations induced by the proposed triplet relationships, each of which depicts a hyper-similarity among three samples. Both theoretical analysis and experimental results demonstrate the effectiveness of the proposed method.

Methodology

Notations and Problem Formulation

Given the data matrix $\boldsymbol{X} = \{\boldsymbol{x}_i \in \mathbb{R}^D\}_{i=1}^N$, subspace clustering (SC) algorithms solve the optimization problem in (1) to generate the coefficient matrix $\boldsymbol{C}$, of which each entry c_{ij} reflects the similarity of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. For each data sample $\boldsymbol{x}_j \in \boldsymbol{\mathcal{S}}$, the SC algorithms take advantage of the self-expressive property, *i.e.*, each data sample can be reconstructed by a linear combination of other points in the dataset (Elhamifar and Vidal 2013; Belkin and Niyogi 2001). As a result, $\boldsymbol{x}_j$ can be written as

$$\boldsymbol{x}_j = \boldsymbol{X}\boldsymbol{c}_j, \ s.t. \ c_{jj} = 0, \tag{2}$$

where the data matrix $\boldsymbol{X}$ is considered as a self-expression dictionary and $\boldsymbol{c}_j = [c_{1j}, c_{2j}, \cdots, c_{Nj}]$ contains the coefficients of the combination. Considering a regularization with various well-designed norms on $\boldsymbol{c}_j$, the calculation system in (2) has the ability to preserve only the combinations among samples in $\boldsymbol{\mathcal{S}}$. Inspired by (Peng et al. 2016), we first collect the nearest neighbors for $\boldsymbol{x}_j$ with top m coefficients in $\boldsymbol{c}_j$.

Definition 1. (m Nearest Neighbors) *The m nearest neighbors for each data point $\boldsymbol{x}_j$,* i.e. *$\boldsymbol{N}_m(\boldsymbol{x}_j) \in \mathbb{R}^{1\times m}$, are defined as follows:*

$$\boldsymbol{N}_m(\boldsymbol{x}_j) = \arg\max_{\{\boldsymbol{x}_{\boldsymbol{i}_l}\}} \sum_{l=1}^{m} |c_{\boldsymbol{i}_l, j}|, \tag{3}$$

where $\boldsymbol{i}_l$ is the set of ordinals for the nearest neighbors, $c_{\boldsymbol{i}_l,j}$ denotes the coefficient of $\boldsymbol{x}_{\boldsymbol{i}_l}$ and $\boldsymbol{x}_j$.

Based on the generated nearest neighbors, we define the triplet relationship to explore the local hyper-correlation among samples.

Definition 2. (Triplet Relationship) *The three samples $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ form a triplet relationship if and only if they satisfy:*

$$\mathbf{1}_{\boldsymbol{x}_i \in \boldsymbol{N}_m(\boldsymbol{x}_j)} \times \mathbf{1}_{\boldsymbol{x}_j \in \boldsymbol{N}_m(\boldsymbol{x}_k)} \times \mathbf{1}_{\boldsymbol{x}_k \in \boldsymbol{N}_m(\boldsymbol{x}_i)} = 1, \tag{4}$$

where $\mathbf{1}_{\boldsymbol{x}\in\boldsymbol{N}_m}$ is the indicator function which equals to 1 if $\boldsymbol{x} \in \boldsymbol{N}_m$ and 0 otherwise.

For easy illustration, we introduce the triplet matrix $\boldsymbol{T} \in \mathbb{R}^{n\times 3}$, where n denotes the number of triplets. Each row of $\boldsymbol{T}$, *i.e.*, $\boldsymbol{\tau} = \{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k\}$, denotes a triplet where $\boldsymbol{x}_i, \boldsymbol{x}_j$ and $\boldsymbol{x}_k$ satisfy the requirement in (4).

Proposition 1. *Given arbitrary three samples in one triplet,* i.e., *$\boldsymbol{\tau} = \{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k\} \in \boldsymbol{T}$, we have $\{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k\} \subset S$, where $\boldsymbol{T}$ denotes the set of triplets and S denotes a specific subspace.*

Proof. Let $\widetilde{\boldsymbol{c}}$ be the optimal solution of the optimization function:

$$\min_{\boldsymbol{c}} \|\boldsymbol{x} - \boldsymbol{X}\boldsymbol{c}\|_2^2 + \lambda\|\boldsymbol{c}\|_\xi, \tag{5}$$

where $\|\boldsymbol{c}\|_\xi$ denotes the ℓ_ξ norm of $\boldsymbol{c}$. We divide the vector $\widetilde{\boldsymbol{c}}$ into the following two parts:

$$\widetilde{\boldsymbol{c}} = \left[\widetilde{\boldsymbol{c}}_{\boldsymbol{D}_{\boldsymbol{x}}}, \widetilde{\boldsymbol{c}}_{\boldsymbol{D}_{-\boldsymbol{x}}}\right]^\top, \tag{6}$$

where $\boldsymbol{D}_{\boldsymbol{x}}$ and $\boldsymbol{D}_{-\boldsymbol{x}}$ denote the collections of intra-cluster samples and inter-cluster samples of the sample $\boldsymbol{x}$, respectively. According to the Intra-subspace Projection Dominance (IPD) which is proved in (Peng et al. 2016), we have

$$\left[\widetilde{\boldsymbol{c}}_{\boldsymbol{D}_{\boldsymbol{x}}}\right]_{r_{\boldsymbol{x}},1} > \left[\widetilde{\boldsymbol{c}}_{\boldsymbol{D}_{-\boldsymbol{x}}}\right]_{1,1}, \tag{7}$$

for ℓ_ξ being equal to ℓ_1, ℓ_2, ℓ_∞ and nuclear norm, where $\left[\widetilde{\boldsymbol{c}}_{\boldsymbol{D}_{\boldsymbol{x}}}\right]_{r_{\boldsymbol{x}},1}$ denotes the $r_{\boldsymbol{x}}$-th largest absolute value of the entries of $\left[\widetilde{\boldsymbol{c}}_{\boldsymbol{D}_{\boldsymbol{x}}}\right]$, $r_{\boldsymbol{x}}$ denotes the dimensionality of the subspace of the intra-cluster samples of $\boldsymbol{x}$.

Assume a binary coefficient matrix $\boldsymbol{C}^*$ satisfies the following function:

$$c_{ij}^* = \begin{cases} 1, & c_{ij} \in \boldsymbol{c}_{r_c,m}, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $\boldsymbol{c}_{r_c,m}$ denotes the m largest values of $\boldsymbol{c}$. According to (4), the triplet $\boldsymbol{\tau} = \{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k\}$ satisfies $c_{ij}^* \times c_{jk}^* \times c_{ki}^* = 1$. Besides, we have $m \ll \frac{N}{K}$, hence the preserved m values are of the intra-cluster samples of $\boldsymbol{x}$. Since $c_{ij}^* \times c_{jk}^* \times c_{ki}^* = 1$ and c^* can only take the value of 0 and 1, we have $c_{ij}^* = c_{jk}^* = c_{ki}^* = 1$. Therefore, the three samples belong to same subspace, *i.e.*, $\{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k\} \subset S$. □

In contrast to the traditional pairwise relationship, the triplet is more robust for partitioning the inter-cluster samples near the intersection of two sub-manifolds due to the complementarity derived from (4). Meanwhile, the triplet relationship depicts a local geometrical structure which enables us to jointly estimate the K and clustering using a greedy framework according to the local density among triplets.

In the clustering process, we iteratively fuse samples into the clusters, *i.e.*, $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$, where $\widehat{K}$ denotes the estimated number of clusters. Therefore, In the I-th iteration, we define the set of "in-clusters" triplets $\boldsymbol{T}_{in}^I$ which are already assigned into clusters, and the set of "out-of-clusters" triplets $\boldsymbol{T}_{out}^I$ which should be assigned in the following iterations. For easy illustration, we reshape the matrices $\boldsymbol{T}_{in}^I \in \mathbb{R}^{p\times 3}$ and $\boldsymbol{T}_{out}^I \in \mathbb{R}^{q\times 3}$ to the vectors $\boldsymbol{X}_{in}^I \in \mathbb{R}^{3p}$ and $\boldsymbol{X}_{out}^I \in \mathbb{R}^{3q}$, both of which preserve the frequency of each sample. We then propose to optimize two new rewards, *i.e.*, the model selection and the fusion reward.

Definition 3. (Model Selection Reward) *The model selection reward $R_m(\mathcal{C})$ for the clusters $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$ is defined as:*

$$R_m(\mathcal{C}) = \sum_i f(\mathcal{C}_i|\boldsymbol{X}_{out}^I) - \lambda_m \sum_i f(\mathcal{C}_i|\boldsymbol{X}_{in}^I), \tag{9}$$

where $f(\mathcal{C}|\boldsymbol{X})$ is a counting function on the frequency that $\boldsymbol{x} \in \boldsymbol{X}_{out}^I$ or $\boldsymbol{x} \in \boldsymbol{X}_{in}^I$ for all $\boldsymbol{x} \in \mathcal{C}_i$, λ_m denotes the trade-off.

By maximizing the model selection reward $R_m(\mathcal{C})$, we generate the initialized cluster $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$ with the following two advantages where $\widehat{K}$ is the estimated number of clusters. 1) Samples in $\mathcal{C}$ are of high density, *i.e.*, have large amount of correlated samples in $\boldsymbol{X}_{out}$, which enables to merge as many in the next iteration; 2) Each $\mathcal{C}$ is of little correlation with samples in $\boldsymbol{X}_{in}$, which eliminates the overlap of the inter-cluster. Thus, we can simultaneously estimated $\widehat{K}$ and initialize the clusters by optimizing the model selection reward R_m.

Definition 4. (Fusion Reward) *The fusion reward optimizes the probability that $\boldsymbol{x}_j \in \boldsymbol{X}_{out}$ being assigned into the cluster $\mathcal{C}_i$, which is defined as:*

$$R_f^i(\mathcal{C}_i|\boldsymbol{x}_j \in \boldsymbol{X}_{out}) = f(\boldsymbol{x}_j|\mathcal{C}_i) + \lambda_f f(\boldsymbol{N}_m(\boldsymbol{x}_j)|\boldsymbol{N}_m(\mathcal{C}_i)), \tag{10}$$

where $\boldsymbol{N}_m(\boldsymbol{x}_j)$ denotes the m nearest neighbors of $\boldsymbol{x}_j$ and $\boldsymbol{N}_m(\mathcal{C}_i)$ denotes the set of m nearest neighbors of samples in $\mathcal{C}_i$, λ_f denotes the trade-off.

We calculate $\widehat{K}$ fusion rewards $\{R_f^i\}_{i=1}^{\widehat{K}}$ for each $\boldsymbol{x}_j$, which represent the probabilities that $\boldsymbol{x}_j$ being assigned into clusters $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$, respectively. We then merge $\boldsymbol{x}_j$ into the cluster with the largest fusion reward, and move the $\boldsymbol{x}_j$ from $\boldsymbol{X}_{out}$ to $\boldsymbol{X}_{in}$.

To determine the first triplet for construct a new cluster, we propose to maximize the local density defined as follows.

Definition 5. (Local Density) *The local density of the triplet $\boldsymbol{\tau}$ regarding to the $\boldsymbol{X}_{out}$ is defined as follows:*

$$\rho(\boldsymbol{\tau}, \boldsymbol{X}_{out}) = \sum_{j=1}^{|\boldsymbol{n}|} f(\boldsymbol{x}_{\boldsymbol{n}_j}|\boldsymbol{X}_{out}), \tag{11}$$

where $\boldsymbol{x}_{\boldsymbol{n}_j}$ denotes the sample in current triplet $\boldsymbol{\tau}$ and $\boldsymbol{n}$ is the set of their ordinals, $|\boldsymbol{n}|$ denotes the scale of $\boldsymbol{n}$.

Also, to determine the optimal triplet to merge into the initialized clusters, we define the connection score as follows.

Algorithm 1 : Automatic Subspace Clustering (autoSC)

Input: $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{D\times N}$.

Calculate the correlation matrix $\boldsymbol{C}$ by (1);
for $i = 1 : N$ **do**
  Calculate the m Nearest Neighbors $\boldsymbol{N}_m(\boldsymbol{x}_i)$ by (3);
end for
Generate the triplet matrix $\boldsymbol{T} \in \mathbb{R}^{n\times 3}$ by (14);
Reshape $\boldsymbol{T}$ to $\boldsymbol{X}_{out} \in \mathbb{R}^{3n}$, $\boldsymbol{X}_{in} = \varnothing$;
$\widetilde{K} = 1$;
Calculate $\boldsymbol{\tau}_{ini}^{\widetilde{K}}$ by (15);
while $\rho(\boldsymbol{\tau}_{ini}^{\widetilde{K}}, \boldsymbol{X}_{out}) > \rho(\boldsymbol{\tau}_{ini}^{\widetilde{K}}, \boldsymbol{X}_{in})$ **do**
  $\mathcal{C}_{\widetilde{K}} = \boldsymbol{\tau}_{ini}^{\widetilde{K}}$;
  $\mathcal{C}_{\widetilde{K}} = \mathcal{C}_{\widetilde{K}} \cup \{\boldsymbol{\tau}^*\}$ where $\boldsymbol{\tau}^*$ is calculated by (16);
  $\boldsymbol{X}_{in} = \boldsymbol{X}_{in} \cup \boldsymbol{\tau}_{ini}^{\widetilde{K}} \cup \{\boldsymbol{\tau}^*\}$;
  $\boldsymbol{X}_{out} = \boldsymbol{X}_{out}/(\boldsymbol{\tau}_{ini}^{\widetilde{K}} \cup \{\boldsymbol{\tau}^*\})$;
  $\widetilde{K} = \widetilde{K} + 1$;
  Calculate $\boldsymbol{\tau}_{ini}^{\widetilde{K}}$ by (15);
end while
Merge $\mathcal{C}_i$ and $\mathcal{C}_j$ if we have (18); Get $\widehat{K}$ clusters;
for $j = 1 : |\boldsymbol{X}_{out}|$ **do**
  Calculate $\mathcal{C}^*$ for $\boldsymbol{x}_j$ by (19);
end for

Output: The cluster assignment $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$.

Definition 6. (Connection Score) *The connection score of the sample $\boldsymbol{x}_i$ towards the sample $\boldsymbol{x}_j$ is defined as:*

$$s(\boldsymbol{x}_i, \boldsymbol{x}_j) = f\left(\boldsymbol{x}_i \middle| \bigcap_{k=1}^{n'} \left(\mathbf{1}_{\boldsymbol{x}_j\in\boldsymbol{\tau}_k} \times \boldsymbol{\tau}_k\right)\right), \tag{12}$$

where $\mathbf{1}_{\boldsymbol{x}_j\in\boldsymbol{\tau}_k}$ is equal to 0 when $\boldsymbol{x}_j \in \boldsymbol{\tau}_k$ and otherwise equal to 1, n' is the number of all triplets in $\boldsymbol{T}_{out}$.

Note both definitions can be easily extended with different components, *e.g.*, $\boldsymbol{\tau}$'s or $\mathcal{C}$'s, of which the extension can be found in the supplementary material.

The Clustering Model based on Triplets

We greedily optimize the proposed model selection reward R_m and the fusion reward R_f in autoSC to simultaneously estimate the number of clusters and generate the segmentation among samples:

$$\begin{aligned} &\max_{\boldsymbol{\mathcal{G}}, \widehat{K}} \sum_{k=1}^{\widehat{K}} R_m(\mathcal{G}_k) + \lambda \sum_{k=1}^{\widehat{K}} R_f(\mathcal{G}_k|\boldsymbol{X}), \\ &s.t.\ \mathcal{G}_k \bigcap \mathcal{G}_{k'\neq k} = \varnothing, \bigcup_{k=1}^{\widehat{K}} \mathcal{G}_k = [1, \cdots, N], \end{aligned} \tag{13}$$

where λ is the trade-off, $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_1, \cdots, \mathcal{G}_{\widehat{K}}\}$ denotes the set of the result groups, $\widehat{K}$ is the estimated number of clusters and $[1, \cdots, N]$ denotes the universal ordinal set of samples.

We illustrate the proposed autoSC in Algorithm 1. In this section, we interpret the implementation details of the optimization in three steps including: 1) generating the triplet

relationships $\boldsymbol{T}$ from the coefficient matrix $\boldsymbol{C}$; 2) estimating the number of clusters $\widehat{K}$ and initializing the clusters $\mathcal{C}$; 3) assigning the samples $\boldsymbol{x} \in \boldsymbol{X}_{out}$ into proper cluster.

The Generation of Triplets The coefficient matrix $\boldsymbol{C}$ reflects the correlations among samples (Elhamifar and Vidal 2009). Larger value indicates stronger belief for the connection between samples. For instance, $c_{ij} > c_{ik}$ indicates a larger probability for $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ being in the same cluster over $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$. Accordingly, we explore the intrinsic local correlations among samples by the proposed triplets derived from $\boldsymbol{C}$.

Many subspace representations guarantee the mapping invariance via a dense coefficient matrix $\boldsymbol{C}$. However, the generation of triplets relies only on the strongest connections to avoid the wrong assignment. Therefore, for each column of $\boldsymbol{C}$, *i.e.*, $\boldsymbol{c}_i$, we preserve only the top m values which are then modified to 1 for a new binary coefficient matrix $\boldsymbol{C}^*$.

Then, we extract each triplet from $\boldsymbol{C}^*$ by the following function:

$$\begin{aligned} &\boldsymbol{\tau} = \{\boldsymbol{x}_{\boldsymbol{n}_1}, \boldsymbol{x}_{\boldsymbol{n}_2}, \boldsymbol{x}_{\boldsymbol{n}_3}\} \in \boldsymbol{T}, \\ &\text{if and only if}: \quad c^*_{\boldsymbol{n}_1\boldsymbol{n}_2} \times c^*_{\boldsymbol{n}_2\boldsymbol{n}_3} \times c^*_{\boldsymbol{n}_3\boldsymbol{n}_1} = 1, \end{aligned} \tag{14}$$

where c^*_{xy} denotes the xy-th value of $\boldsymbol{C}^*$. Note each sample $\boldsymbol{x}$ can belong to many triplets. Therefore, we consider each $\boldsymbol{\tau}$ as a meta-element in the clustering, which improves the robustness due to the complementarity constraints.

The Initialization of the Clusters In the I-th iteration, we first determine the initialized triplet (termed as $\boldsymbol{\tau}^I_{ini}$) from $\boldsymbol{T}_{out}$ to be the basement of the cluster $\mathcal{C}$. Then, we merge the most correlated samples of $\boldsymbol{\tau}^I_{ini}$ into $\mathcal{C}$. Finally, we detect the repetitiveness between $\mathcal{C}$ and the "in-cluster" samples $\boldsymbol{X}^I_{in}$ to avoid the redundancy.

Following (Rodriguez and Laio 2014), we initialize a new cluster from $\boldsymbol{\tau}^I_{ini}$ with highest local density:

$$\boldsymbol{\tau}^I_{ini} = \arg\max_{\boldsymbol{\tau}} \ \rho(\boldsymbol{\tau}, \boldsymbol{X}^I_{out}), \tag{15}$$

where ρ calculate the local density as shown in Definition 5. The high local density of the triplet evokes the most connections between $\boldsymbol{\tau}_i$ and other triplets, which induces the most connections between $\boldsymbol{x}_{\boldsymbol{n}^i_j}$ and other samples in $\boldsymbol{X}^I_{out}$.

Once the initialized triplet $\boldsymbol{\tau}^I_{ini}$ is determined, we iteratively extend the cluster $\mathcal{C}$ by fusing the most confident triplets. For each triplet $\boldsymbol{\tau}_i$ in $\boldsymbol{T}_{out}$, we calculate the sum of the connection score regarding the samples in $\mathcal{C}$ to greedily determine whether the samples in $\boldsymbol{\tau}_i$ should be assigned into $\mathcal{C}$:

$$\begin{aligned} &\boldsymbol{\tau}^* = \arg\max_{\boldsymbol{\tau}} \sum_{j=1}^{3}\sum_{\kappa}^{|\boldsymbol{m}|} s_{\boldsymbol{n}_j\boldsymbol{m}_\kappa}, \\ &s.t. \sum_{j=1}^{3}\sum_{\kappa}^{|\boldsymbol{m}|} s_{\boldsymbol{n}_j\boldsymbol{m}_\kappa} > 1; \{\boldsymbol{x}_{\boldsymbol{n}_j}\}_{j=1}^{3} \in \boldsymbol{\tau}; \{\boldsymbol{x}_{\boldsymbol{m}_\kappa}\}_{\kappa=1}^{|\boldsymbol{m}|} \in \mathcal{C}, \end{aligned} \tag{16}$$

where $\boldsymbol{n}, \boldsymbol{m}$ denote the set of ordinals for the samples in $\boldsymbol{\tau}$ and $\mathcal{C}$, respectively. We iteratively update the auxiliary sets $\boldsymbol{T}^I_{out}$, $\boldsymbol{T}^I_{in}$, $\boldsymbol{X}^I_{out}$ and $\boldsymbol{X}^I_{in}$ in the iteration.

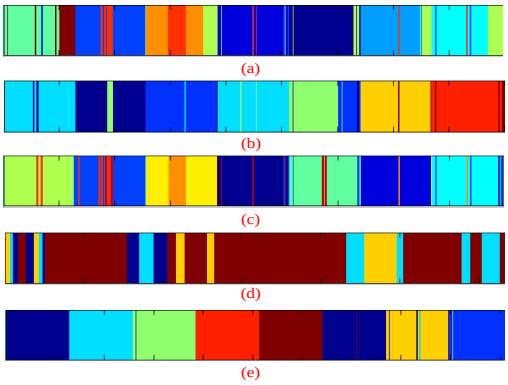

Figure 1: The visualization of the clustering labels for four methods, *i.e.*, (a) SCAMS, (b) DP, (c) SVD, (d) DP-space and (e) the proposed autoSC. The experiments are executed on the extended Yale B dataset with 8 subjects. As shown, SCAMS over-segments the samples, while DP-space assigns the majority into one cluster. The proposed autoSC does not suffer from these problems.

We terminate the process of model selection and get $\widetilde{K}$ clusters if and only if $\boldsymbol{\tau}^{\widetilde{K}+1}_{ini}$ satisfies:

$$\rho(\boldsymbol{\tau}^{\widetilde{K}+1}_{ini}, \boldsymbol{X}^{\widetilde{K}+1}_{out}) \leq \rho(\boldsymbol{\tau}^{\widetilde{K}+1}_{ini}, \boldsymbol{X}^{\widetilde{K}+1}_{in}). \tag{17}$$

Specifically, if the samples of $\boldsymbol{\tau}^{\widetilde{K}+1}_{ini}$ are of high frequency in $\boldsymbol{X}^{\widetilde{K}+1}_{in}$, *i.e.*, the triplet with the highest local density in $\boldsymbol{T}^{\widetilde{K}+1}_{out}$ is already contained in $\boldsymbol{T}^{\widetilde{K}+1}_{in}$, we consider the clusters are sufficient for modeling the intrinsic sub-manifolds.

We also introduce an alternative step to check the redundancy among clusters $\{\mathcal{C}_i\}_{i=1}^{\widetilde{K}}$. We calculate the connection scores s for small-scale clusters against others, and merge the highly correlated clusters $\mathcal{C}_i$ and $\mathcal{C}_j$ if we have

$$s_{ij} > \min(|\mathcal{C}_i|, |\mathcal{C}_j|), \tag{18}$$

where $|\mathcal{C}|$ denotes the number of samples in $\mathcal{C}$. We then get the initialized clusters $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$, where $\widehat{K}$ is the estimated number of clusters and $\widehat{K} \leq \widetilde{K}$.

The Assignment of the Remaining Samples In this stage, we assign each of the remaining samples into clusters which evokes an optimal fusion reward. For $\boldsymbol{x}_j$, we find its optimal cluster $\mathcal{C}^*$ by the following equation:

$$\mathcal{C}^* = \arg\max_{\mathcal{C}_i} \ R_f(\mathcal{C}_i|\boldsymbol{x}_j), \ \ i \in \{1, 2, \cdots, \widehat{K}\}, \tag{19}$$

where $R_f(\mathcal{C}|\boldsymbol{x})$ is the fusion reward defined by (10).

Experiment

Setup

In the experiments, we first compare the proposed autoSC with various automatic methods on two benchmark datasets, *i.e.*, the extended Yale B and the COIL-20 dataset. Then, we verify the robustness of the proposed method with combinations to different $\boldsymbol{C}$ derived from various subspace representations. We also design comprehensive evaluation metrics to

Table 1: Overall Comparison between autoSC and other algorithms on subsets of the extended Yale B and COIL-20 dataset. The coefficient matrix $\boldsymbol{C}$ derived from SMR is utilized as the correlation matrix of DP, SVD and the proposed autoSC. As shown, autoSC achieves the state-of-the-art performance on all reported configurations.

Methods	Metrics	extended Yale B			COIL-20		
		8	15	30	5	10	15
SCAMS	NC_e	9.26	23.60	76.22	8.48	19.72	32.40
	NMI	0.7183	0.7272	0.7266	0.5885	0.6527	0.6668
DP	NC_e	3.06	7.84	24.76	2.22	5.30	9.72
	NMI	0.6196	0.5026	0.2166	0.6864	0.4467	0.3643
SVD	NC_e	2.40	9.06	24.00	0.48	2.58	8.36
	NMI	0.7078	0.4993	0.2808	0.7024	0.7127	0.7224
DP-space	NC_e	2.08	8.96	23.92	0.78	4.78	9.38
	NMI	0.0343	0.0226	0.0406	0.0904	0.0829	0.0718
autoSC	NC_e	**0.76**	**2.08**	**4.98**	**0.38**	**1.18**	**0.80**
	NMI	**0.9062**	**0.8589**	**0.8287**	**0.8315**	**0.7701**	**0.7266**

validate the clustering performance, *e.g.*, the error rate of the number of clusters and the triplets, *etc*. All reported results are the average of 50 trials.

Methods We make comparisons with the following methods: SCAMS (Li et al. 2016), density peak based method (DP) (Rodriguez and Laio 2014), singular value decomposition based method (SVD) (Liu et al. 2013) and DP-space (Wang and Zhu 2015). Besides, we utilize the following subspace representation methods to generate different coefficient matrix $\boldsymbol{C}$, including LRR (Liu, Latecki, and Yan 2010), CASS (Lu et al. 2015), LSR (Lu et al. 2012), smooth representation (SMR) (Hu et al. 2014) and ORGEN (You et al. 2016). The coefficient matrix $\boldsymbol{C}$ is then used to calculate the triplet relationships for the proposed autoSC.

Metrics To evaluate the performance of the proposed triplets, we define the error rate $\mathcal{A}$ as follows:

$$\mathcal{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{3 - f(\boldsymbol{\tau}_i|\boldsymbol{g}_i^*)}{2}, \tag{20}$$

where n denotes the number of the triplets and $f(\boldsymbol{\tau}|\boldsymbol{g}^*)$ is the counting function on the frequency that $\boldsymbol{x} \in \boldsymbol{g}^*$ for all $\boldsymbol{x} \in \boldsymbol{\tau}$. Here the output of f ranges from 0 to 2. The dynamic set $\boldsymbol{g}_i^*$ consists of samples in one subspace S according to the ground truth, where S contains as many samples in $\boldsymbol{\tau}_i$ as possible.

We introduce the error rate of the number of clusters (NC_e) as the primary evaluation metric for the clustering methods which estimate the number of clusters $\widehat{K}$ automatically:

$$\text{NC}_e = \frac{1}{M}\sum_{i=1}^{M}|\widehat{K}_i - K|, \tag{21}$$

where K is the real number of clusters, M is the number of trials and $\widehat{K}_i$ is the estimated number of clusters in the i-th trial. We also use the standard normalized mutual information (NMI) (Li et al. 2003) to measure the similarity between two clustering distributions, *i.e.*, the prediction and the ground truth. With respect to NMI, the entropy illustrates the non-determinacy of one clustering to the other, and the mutual information quantifies the amount of information that one variable obtains from the other.

Table 2: The evaluation results of $\mathcal{A}$ for the proposed autoSC on the extended Yale B (eYaleB) and COIL-20 dataset. For each column, we utilize the coefficient matrix of a specific subspace representation module. As shown, the autoSC achieves consistency on the calculation of the triplets, which guarantees the performance of the model selection and clustering process.

Datasets		LRR	CASS	LSR	SMR	ORGEN
eYaleB	8	0.0155	0.0158	0.0144	0.0135	0.0166
	15	0.0147	0.0148	0.0148	0.0149	0.0145
	30	0.0176	0.0157	0.0181	0.0181	0.0177
COIL20	5	0.0185	0.0195	0.0188	0.0175	0.0196
	10	0.0252	0.0198	0.0188	0.0182	0.0210
	15	0.0224	0.0203	0.0212	0.0196	0.0215

The Comparisons among Automatic Clustering

We conduct experiments on two datasets (the extended Yale B and the COIL-20) with different number of subjects, and compare four methods with the proposed autoSC on the metrics of NC_e and NMI. For DP, SVD and our autoSC, the optimization module in SMR (Hu et al. 2014) is employed to generate the coefficient matrix $\boldsymbol{C}$. The DP-space method simultaneously estimates the $\widehat{K}$ and finds the subspaces without the requirement of coefficient matrix. All parameters of the contrasted methods are tuned to be the best. Table 1 and Figure 1 show the performance.

As shown, the averaged NC_e of autoSC is relatively smaller than others on all experimental configurations, meaning that we can give a close estimation on the number of clusters. For example, the estimated $\widehat{K}$ on the extended Yale B with 8 subjects has the deviation which is less than 1, and evokes the NMI higher than 0.9. SVD achieves comparable result on the small-scale configuration of each dataset, but the performance turns to be bad when the number of samples increases. It is mainly because that the largest gap between

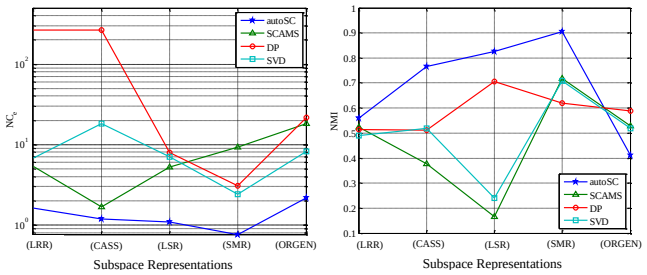

Figure 2: The extension to different subspace representations on the extended Yale B dataset with 8 subjects. The left figure denotes the comparison of four methods using the NC_e metric while the right one is regarding the NMI. Each point in the curves is derived by the combination of clustering method and subspace representation. As shown, the proposed autoSC achieves consistent performance on the evaluation of NC_e.

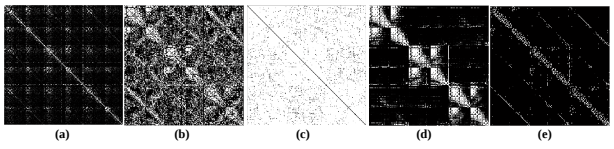

Figure 3: The coefficient matrix $\boldsymbol{C}$ derived from different methods on the extended Yale B dataset with 3 subjects. The white regions denote the locations with non-zero coefficients. Different methods, *i.e.*, (a) LRR, (b) CASS, (c) LSR, (d) SMR and (e) ORGEN, induce $\boldsymbol{C}$ with different characteristics, *e.g.*, (d) derived from SMR is block-diagonal yet (e) derived from ORGEN is sparse.

the pair of singulars decreases when the number of clusters getting larger. SCAMS performs comparably for the NMI on both datasets, however, as is illustrated in Figure 1 (a) and Table 1, it provides a much larger $\widehat{K}$ than the ground truth, *e.g.*, $\widehat{K} > 100$ when $K = 30$ on the extended Yale B dataset. The NMI is friendly to the situation of over-segmentation, making the metric NC_e be the primary evaluation of the SCAMS method. The DP-space performs well on NC_e, but has bad performance on the NMI. It is because that most samples are assigned into one cluster, and the other clusters are of small scale. As a contrast, our autoSC achieves favorable results with a satisfied generalization ability.

The Robustness to Subspace Representations

As illustrated, the methods including SCAMS (Li et al. 2016), SVD (Liu et al. 2013) and our autoSC require the coefficient matrix $\boldsymbol{C}$ as input. Also, for DP (Rodriguez and Laio 2014), it needs to calculate the distance among samples. We calculate the distance d_{ij} between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ by $d_{ij} = \frac{1}{c_{ij}}$ rather than the simple Euclidean distance. To verify the robustness of the proposed autoSC regarding various subspace representations, we calculate the coefficient matrix $\boldsymbol{C}$ using 5 subspace representation modules, followed by the combinations with the 4 methods which automatically estimate the number of clusters and segment the samples.

Table 2 shows the evaluation results of $\mathcal{A}$ on both datasets with the combinations of 5 subspace representations. The experimental results on the extended Yale B dataset with 8 subjects are reported in Figure 2, while other similar experiments with different configurations can be found in the supplementary material. Moreover, we visualize the coefficient matrix $\boldsymbol{C}$ derived from 5 subspace representation modules in Figure 3. We can see from Figure 2 that the SCAMS, DP and SVD methods are sensitive on the choice of the subspace representation module. For example, DP estimates the $\widehat{K}$ as a relatively close value to the ground truth when combined with SMR ($\text{NC}_e = 3.06$), but generates a totally wrong estimation when combined with LRR ($\text{NC}_e = 265.60$). Different subspace representation modules generate coefficient matrices with various intrinsic properties (Vidal 2011), thus the parameter for truncation error ϵ needs to be tuned carefully.

For the proposed autoSC, it is stable on different combinations considering the metric of NC_e and $\mathcal{A}$, which demonstrates the complementary ability of the proposed method. For all combinations, the error rate of the triplets obtained from (14) is less than 2%, which guarantees the consistency of the proposed autoSC with different kind of $\boldsymbol{C}$. Furthermore, it shows better performance when combined with CASS, LSR and SMR than other combinations on both metrics in Figure 2. The reason lies on the guarantee of the mapping invariance which is termed as the grouping effect (Lu et al. 2015; Lu et al. 2012; Hu et al. 2014), together with the filtering of weak connections and the self-constraint among samples within triplets. As shown in Figure 3 (b), (c), (d), the coefficient matrices are dense while it shows block-diagonal structure in Figure 3 (d) and each block corresponds to one cluster. Therefore, the nearest neighbors which are used to generate the triplets can be chosen precisely. As shown in Figure 3 (b), (c), (d), the coefficient matrices are dense and block-diagonal which evokes effective triplets. The autoSC can not achieve satisfied performance when combined with the ORGEN as shown in Figure 2. It is because that the $\boldsymbol{C}$ derived from ORGEN is too sparse which contains insufficient localities for constructing effective triplets.

Conclusion

In this paper, we propose the autoSC method to simultaneously estimate the number of clusters and segment the samples according to the coefficient matrix derived by off-the-shelf subspace representations. We consider this problem as an optimization process on two reward functions at the same time, of which the model selection reward constrains the number of clusters and the fusion reward facilitates the segmentation of the samples. These two functions are greedily maximized during the clustering process, utilizing the predefined data structure termed triplet relationship. The triplet is more robust than the pairwise relationships when partitioning the inter-cluster samples near the intersection of two sub-manifolds due to the complementarity of multiple constraints. Besides, it evokes mutual restrictions of neighbors thus depicts a local geometrical structure, by which we can calculate the segmentation greedily. Extensive experiments on the benchmarks demonstrate the effectiveness of our approach.

Acknowledgments

This research was sponsored by NSFC (61620106008, 61572264), CAST (YESS20150117), Huawei Innovation Research Program (HIRP), and IBM Global SUR award.

References

[Beier, Hamprecht, and Kappes 2015] Beier, T.; Hamprecht, F. A.; and Kappes, J. H. 2015. Fusion moves for correlation clustering. In *CVPR*.

[Belkin and Niyogi 2001] Belkin, M., and Niyogi, P. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*.

[Cheng et al. 2016] Cheng, Y.; Wang, Y.; Sznaier, M.; and Camps, O. 2016. Subspace clustering with priors via sparse quadratically constrained quadratic programming. In *CVPR*.

[Elhamifar and Vidal 2009] Elhamifar, E., and Vidal, R. 2009. Sparse subspace clustering. In *CVPR*.

[Elhamifar and Vidal 2013] Elhamifar, E., and Vidal, R. 2013. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Favaro, Vidal, and Ravichandran 2011] Favaro, P.; Vidal, R.; and Ravichandran, A. 2011. A closed form solution to robust subspace estimation and clustering. In *CVPR*.

[Feng et al. 2014] Feng, J.; Lin, Z.; Xu, H.; and Yan, S. 2014. Robust subspace segmentation with block-diagonal prior. In *CVPR*.

[Gao, Tsang, and Chia 2013] Gao, S.; Tsang, I. W.; and Chia, L. T. 2013. Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Guo, Gao, and Li 2013] Guo, Y.; Gao, J.; and Li, F. 2013. Spatial subspace clustering for hyperspectral data segmentation. In *SDIWC*.

[Hu et al. 2014] Hu, H.; Lin, Z.; Feng, J.; and Zhou, J. 2014. Smooth representation clustering. In *CVPR*.

[Kim et al. 2014] Kim, S.; Chang, D. Y.; Nowozin, S.; and Kohli, P. 2014. Image segmentation usinghigher-order correlation clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Li and Vidal 2015] Li, C. G., and Vidal, R. 2015. Structured sparse subspace clustering: A unified optimization framework. In *CVPR*.

[Li et al. 2003] Li, M.; Chen, X.; Li, X.; and Ma, B. 2003. Clustering by compression. In *IEEE International Symposium on Information Theory*.

[Li et al. 2016] Li, Z.; Yang, S.; Cheong, L. F.; and Toh, K. C. 2016. Simultaneous clustering and model selection for tensor affinities. In *CVPR*.

[Li, Cheong, and Zhou 2014] Li, Z.; Cheong, L. F.; and Zhou, S. Z. 2014. SCAMS: Simultaneous clustering and model selection. In *CVPR*.

[Liu et al. 2013] Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Liu, Latecki, and Yan 2010] Liu, H.; Latecki, L.; and Yan, S. 2010. Robust clustering as ensembles of affinity relations. In *NIPS*.

[Lu et al. 2012] Lu, C. Y.; Min, H.; Zhao, Z. Q.; Zhu, L.; Huang, D. S.; and Yan, S. 2012. Robust and efficient subspace segmentation via least squares regression. In *ECCV*.

[Lu et al. 2015] Lu, C.; Feng, J.; Lin, Z.; and Yan, S. 2015. Correlation adaptive subspace segmentation by trace lasso. In *CVPR*.

[Lu et al. 2016] Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2016. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *CVPR*.

[Nasihatkon and Hartley 2011] Nasihatkon, B., and Hartley, R. 2011. Graph connectivity in sparse subspace clustering. In *CVPR*.

[Peng et al. 2016] Peng, X.; Yu, Z.; Yi, Z.; and Tang, H. 2016. Constructing the l2-graph for robust subspace learning and subspace clustering. *IEEE Transactions on Cybernetics*.

[Peng, Zhang, and Yi 2013] Peng, X.; Zhang, L.; and Yi, Z. 2013. Scalable sparse subspace clustering. In *CVPR*.

[Purkait et al. 2014] Purkait, P.; Chin, T. J.; Ackermann, H.; and Suter, D. 2014. Clustering with hypergraphs: The case for large hyperedges. In *ECCV*.

[Rodriguez and Laio 2014] Rodriguez, A., and Laio, A. 2014. Clustering by fast search and find of density peaks. *Science*.

[Schölkopf, Platt, and Hofmann 2006] Schölkopf, B.; Platt, J.; and Hofmann, T. 2006. Learning with hypergraphs: Clustering, classification, and embedding. In *NIPS*.

[Schroff, Kalenichenko, and Philbin 2015] Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *ICCV*.

[Shi and Malik 2000] Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Vidal 2011] Vidal, R. 2011. Subspace clustering. *IEEE Signal Processing Magazine*.

[Wang and Zhu 2015] Wang, Y., and Zhu, J. 2015. Dp-space: Bayesian nonparametric subspace clustering with small-variance asymptotics. In *ICML*.

[Yang et al. 2016] Yang, Y.; Feng, J.; Jojic, N.; Yang, J.; and Huang, T. S. 2016. l_0-sparse subspace clustering. In *ECCV*.

[Yin, Gao, and Lin 2016] Yin, M.; Gao, J.; and Lin, Z. 2016. Laplacian regularized low-rank representation and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[You et al. 2016] You, C.; Li, C. G.; Robinson, D. P.; and Vidal, R. 2016. Oracle based active set algorithm for scalable elastic net subspace clustering. In *CVPR*.

[Zhao et al. 2017] Zhao, S.; Yao, H.; Gao, Y.; Ji, R.; and Ding, G. 2017. Continuous probability distribution prediction of image emotions via multitask shared sparse regression. *IEEE Transactions on Multimedia*.