Double Nuclear Norm based Low Rank Representation on Grassmann Manifolds for Clustering

Xinglin Piao[1], Yongli Hu[2,*], Junbin Gao[3], Yanfeng Sun[2] and Baocai Yin[1,2,4]

[1]Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, China
[2]Beijing Key Laboratory of Multimedia and Intelligent Software Technology,
Faculty of Information Technology, Beijing University of Technology, China
[3]Discipline of Business Analytics, The University of Sydney Business School,
The University of Sydney, Camperdown NSW 2006, Australia
[4]Peng Cheng Laboratory, China

piaoxinglin@dlut.edu.cn; huyongli@bjut.edu.cn; junbin.gao@sydney.edu.au;
yfsun@bjut.edu.cn; ybc@bjut.edu.cn

Abstract

Unsupervised clustering for high-dimension data (such as imageset or video) is a hard issue in data processing and data mining area since these data always lie on a manifold (such as Grassmann manifold). Inspired of Low Rank representation theory, researchers proposed a series of effective clustering methods for high-dimension data with non-linear metric. However, most of these methods adopt the traditional single nuclear norm as the relaxation of the rank function, which would lead to suboptimal solution deviated from the original one. In this paper, we propose a new low rank model for high-dimension data clustering task on Grassmann manifold based on the Double Nuclear norm which is used to better approximate the rank minimization of matrix. Further, to consider the inner geometry or structure of data space, we integrated the adaptive Laplacian regularization to construct the local relationship of data samples. The proposed models have been assessed on several public datasets for imageset clustering. The experimental results show that the proposed models outperform the state-of-the-art clustering ones.

1. Introduction

As an active topic in data processing and data mining, data clustering has attracted great interests [2, 13]. A large number of clustering methods [10, 20] have been proposed and successfully used in many applications [41, 17]. In all the clustering methods, the spectral clustering methods [41, 16] based on subspace assumption are considered state-of-the-art methods with promising performance. It is generally assumed that data have intrinsic subspace structures [32] or the data are generated from multiple subspaces. Thus, the datum in a subspace could be linearly represented by a smaller number of other data samples from the same subspace. The key problem of these methods is to obtain a good affinity matrix which usually describes the data similarity determined by the underlying subspace structure. To this end, various subspace clustering methods are proposed. The most representative methods are Sparse Subspace Clustering (SSC) [7] and Low-Rank Representation (LRR) [16], which use sparse and low rank constraints to construct affinity matrix, respectively. Later, researchers utilize the local geometry or structure of the raw data as regularizers such as the Laplacian regularizer, and propose the Non-negative Sparse Laplacian regularized Low Rank Representation (NSLLRR) [40]. From the representation matrix, clustering results can be obtained by using a spectral clustering algorithm such as Normalized Cuts (NCut)[28].

In the aforementioned methods, the data is usually formulated as vectors with Euclidean distance. However, one often faces high dimensional data following nonlinear constraints, especially for imagesets and videos data [39, 34]. Since these high-dimension data are always lie on nonlinear manifold space, the traditional linear methods are no longer valid. For example, an imageset or a video is actually modeled as a data sample on Grassmann manifold. Thus,

*Corresponding author: Yongli Hu (huyongli@bjut.edu.cn)

researchers extended the traditional LRR based methods on the Grassmann manifold for high-dimension data clustering based on the non-Euclidean geometry. The representative methods are Low Rank Representation on Grassmann manifold method (G-LRR)[33], Partial Sum Minimization of Singular Values Representation on Grassmann manifold method (G-PSSVR)[37] and Cascaded Low Rank and Sparse Representation on Grassmann manifold method (G-CLRSR)[35]. Further, combined with the Laplacian regularizer, researchers proposed Laplacian Low-Rank Representation on Grassmann manifold method (G-LLRR)[34, 36] and Laplacian Partial Sum Minimization of Singular Values Representation on Grassmann manifold method (G-LPSSVR)[37] in which authors construct the Laplacian matrix by original data samples.

Although the above methods have achieved good performance in Grassmann manifold clustering, all these methods adopt the traditional nuclear norm as the low rank constraint which would lead to a suboptimal solution [15, 8, 5, 38]. That is because the traditional nuclear norm based low-rank subproblem would tend to over-relaxations of rank components from the representation matrix [42]. In recent works, researchers usually adopt Schatten-p quasi-norm ($0 < p < 1$) instead of the traditional nuclear norm for low rank based problems[22, 27]. Schatten-p norm has decomposable approach and it could construct a more accurate low rank matrix than the traditional nuclear norm [21]. In this paper, we adopt double nuclear norm, a kind of Schatten-p quasi-norm, instead of the traditional single nuclear norm to formulate a new clustering model on Grassmann manifold with non-linear metric for imageset clustering task. We call this model as Double Nuclear norm based Low Rank model on Grassmann manifold (G-DNLR). Further, to better exploit the the local geometrical structure of data space, we introduce the adaptive Laplacian regularizer into the G-DNLR and formulate an adaptive Laplacian regularized G-DNLR which is called adaptive Laplacian Double Nuclear based Low Rank model on Grassmann manifold (G-ALDNLR). The contributions of this paper are following:

- Proposing a new low rank based clustering model on Grassmann manifold for imageset clustering task by utilizing double nuclear norm with non-linear metric;
- Adaptive Laplacian regularizer is introduced into the G-DNLR to formulate G-ALDNLR model for exploiting the local geometrical structure of the data samples;
- An algorithm is proposed to solve the complicated optimization problems involved in the proposed models.

The paper is organized as follows. We introduce the notations and definition of Grassmann manifold in Section 2. Section 3 review the related works. We will introduce the formulation and optimization of the propose G-DNLR and G-ALDNLR models in Section 4 and 5 respectively. Section 6 assesses the clustering performance of the proposed method on several datasets. Finally, conclusions are discussed in Section 7.

2. Notation and Definition of Grasssmann Manifold

2.1. Notation

We use bold lowercase letters for vectors, e.g. $\mathbf{x}, \mathbf{y}, \mathbf{a}$, bold uppercase for matrices, e.g. $\mathbf{X}, \mathbf{Y}, \mathbf{A}$, calligraphy letters for tensors e.g. $\mathcal{X}, \mathcal{Y}, \mathcal{A}$, lowercase letters for scalars such as dimension and class numbers, e.g. m, n, c. $\mathbf{x}_i$ represents the i-th column of matrix $\mathbf{X}$. x_{ij} represents the i-th element in j-th column from matrix $\mathbf{X}$. $\mathbb{R}^{m \times n}$ represents the space of real numbers.

2.2. Definition of Grassmann Manifold

According to [1], the Grassmann manifold consists of all linear p-dimension subspaces in m-dimension Euclidean space $\mathbb{R}^m (0 \le p \le m)$ which is denoted as $\mathcal{G}(p, m)$. Thus, we could construct a Grassmann manifold as below:

$$\mathcal{G}(p, m) = \{\mathbf{Y} \in \mathbb{R}^{m \times p} : \mathbf{Y}^T \mathbf{Y} = \mathbf{I}_p\} / \mathcal{O}(p), \quad (1)$$

where $\mathcal{O}(p)$ represents the p-order orthogonal group. For two Grassmann manifold data samples $\mathbf{Y}_1$ and $\mathbf{Y}_2$, the distance of them could be defined as below [12]:

$$\text{dist}_g(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{1}{2} \|\Pi(\mathbf{Y}_1) - \Pi(\mathbf{Y}_2)\|_F, \quad (2)$$

where $\Pi(\cdot)$ is a mapping function defined as below:

$$\Pi : \mathcal{G}(p, m) \longrightarrow Sym(m), \Pi(\mathbf{Y}) = \mathbf{Y}\mathbf{Y}^T. \quad (3)$$

where $Sym(m)$ represents the space of m-dimension symmetric matrices. With the function $\Pi(\cdot)$, Grassmann manifold could be embedded into the symmetric matrices.

3. Related Works

We first introduce related clustering methods on the Euclidean Space. Given a set of sample vectors $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n] \in \mathbb{R}^{m \times n}$ drawn from a union of c subspaces $\{\mathcal{S}_i\}_{i=1}^c$, where m denotes the dimension of each sample $\mathbf{y}_i$ and n represents the number of samples $\mathbf{Y}$. The task of subspace clustering is to segment the sample set $\mathbf{Y}$ according to the underlying subspaces. By introducing a hypothesized representation matrix $\mathbf{X}$, the data could be self-represented by a linear combination as $\mathbf{Y} = \mathbf{Y}\mathbf{X}$. To avoid the trivial solution, some matrix constraints are adopted on $\mathbf{X}$. In the past decade, sparse and low rank theories have been applied to subspace clustering successfully. Elhamifar and Vidal [7] proposed Sparse Subspace Clustering (SSC) method, which aims to find the sparsest representation matrix $\mathbf{X}$ by using ℓ_1 norm $\|\cdot\|_1$. The SSC model is formulated as follows,

$$\min_{\mathbf{X}} \quad \lambda \|\mathbf{X}\|_1 + \|\mathbf{Y} - \mathbf{Y}\mathbf{X}\|_F^2, \quad (4)$$

where λ is balance parameter and $\|\mathbf{X}\|_1 = \sum_{i=1,j=1}^{n} |x_{ij}|$. Instead of adopting the sparse constraint, Liu *et al.* [16] proposed Low Rank Representation (LRR) method for clustering by using low rank constraint or nuclear norm $\|\cdot\|_*$ on $\mathbf{X}$, which is formulated as follows,

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_* + \|\mathbf{Y} - \mathbf{YX}\|_F^2, \tag{5}$$

where $\|\mathbf{X}\|_* = \sum_i^r \sigma_i(\mathbf{X})$ and $\sigma_i(\mathbf{X})$ represents the i-th singular value of $\mathbf{X}$, r represents the rank of $\mathbf{X}$.

Later, many researchers develop some Laplacian regularizer based subspace clustering methods [19, 24]. The idea of Laplacian regularizer is induced from the graph theory [6]. According to this theory, an undirected local k-connected graph is constructed for $\mathbf{Y}$. This graph could be encoded by a symmetric affinity matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$, where $0 \le w_{ij} \le 1$ reflects the probability that the data points $\mathbf{y}_i$ and $\mathbf{y}_j$ are connected, i.e., $w_{ij} > 0$ means $\mathbf{y}_i$ and $\mathbf{y}_j$ are closer in certain metric or in a local neighbourhood. The local geometry of these data points should be correspondingly reflected in the data representation matrix $\mathbf{X}$, which can be formulated by minimizing the following error,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = 2\mathrm{tr}(\mathbf{X}\mathbf{L}_\mathbf{W}\mathbf{X}^T), \tag{6}$$

where tr$(\cdot)$ represents the trace function of matrix, $\mathbf{L}_\mathbf{W} = \mathbf{D} - \mathbf{W}$ represents the Laplacian matrix, $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix $\mathbf{D}$ with diagonal elements $d_{ii} = \sum_{j=1}^{n} w_{ij}, i = 1, ..., n$. Thus, Yin *et al.* [40] proposed a general Laplacian regularized low-rank representation model by using a graph regularizer called hypergraph as below:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* + \lambda\|\mathbf{X}\|_1 + \alpha\mathrm{tr}(\mathbf{X}\mathbf{L}_\mathbf{W}\mathbf{X}^T) + \beta\|\mathbf{E}\|_1, \\ \text{s.t.} \quad & \mathbf{X} \ge 0, \mathbf{Y} = \mathbf{YX} + \mathbf{E}, \end{aligned} \tag{7}$$

where α, β are balance parameters.

The above related works all construct the representation matrix of data samples by employing Euclidean distance which is not suitable for the high-dimension Grassmann manifold data. Therefore, researchers proposed a series of clustering methods for Grassmann manifold based on the non-distance defined in (2). For a set of Grassmann samples $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$ where $\mathbf{Y}_i \in \mathcal{G}(p, m)$. By generating the (5) on Grassmann, Wang *et al.* [33] proposed a Low Rank model with non-linear metric for Grassmann (G-LRR):

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_* + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_\mathcal{G}, \tag{8}$$

where $\|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_\mathcal{G}$ represents the reconstruction error of the sample $\mathbf{Y}_i$ on Grassmann manifold, $\biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}$ denotes the "combination" of $\{\mathbf{Y}_j\}_{j=1}^{n}$ with the coefficients $\{x_{ji}\}_{i=1,j=1}^{n}$, the symbol $\ominus, \biguplus, \circledast$ are abstract symbols which are used to simulated the "linear" operations on Grassmann manifold. Combined with Laplacian regularizer defined in (6), Wang *et al.*[34, 36] proposed Laplacian Low Rank model on Grassmann manifold (G-LLRR) model:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \lambda\|\mathbf{X}\|_* + \alpha\mathrm{tr}(\mathbf{X}\mathbf{L}_\mathbf{W}\mathbf{X}^T) \\ & + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_\mathcal{G}, \end{aligned} \tag{9}$$

where affinity matrix $\mathbf{W}$ is constructed based on non-linear distance metric defined in (2) based on the raw data $\{\mathbf{Y}_i\}_{i=1}^{n}$. They also proposed a cascaded Low Rank and Sparse model on Grassmann manifold (G-CLRSR) model:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{C}} \quad & \lambda\|\mathbf{X}\|_* + \alpha\|\mathbf{Z}\|_1 + \beta\|\mathbf{X} - \mathbf{XZ}\|_F^2 \\ & + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_\mathcal{G}. \end{aligned} \tag{10}$$

Further, to achieve better low rank representation matrix for clustering, Wang *et al.* [37] adopt Partial Sum Minimization of Singular Values (PSSV) norm to instead the nuclear norm for formulating PSSV Low Rank model on Grassmann manifold (G-PSSVLR) model:

$$\min_{\mathbf{X}} \quad \lambda\|\mathbf{X}\|_{>r} + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_\mathcal{G}, \tag{11}$$

where $\|\cdot\|_{>r}$ represents the PSSV norm defined as below [25]:

$$\|\mathbf{X}\|_{>r} = \sum_{i=r+1}^{n} \sigma_i(\mathbf{X}), \tag{12}$$

where $\sigma_i(\mathbf{X})$ represents the i-th largest singular value of $\mathbf{X}$, r represents the expected rank of $\mathbf{X}$. Similar to G-LLRR, Wang *et al.* also proposed Laplacian G-PSSVLR (G-LPSSVLR) as below [37]:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \lambda\|\mathbf{X}\|_{>r} + \alpha\mathrm{tr}(\mathbf{X}\mathbf{L}_\mathbf{W}\mathbf{X}^T) \\ & + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_\mathcal{G}. \end{aligned} \tag{13}$$

Although the above methods achieve great performance in Grassmann manifold clustering problem, all these methods adopt nuclear norm or PSSV norm which would cause sub-optimal solution of the low rank based problem. Further, the affinity matrices in G-LLRR and G-LPSSVLR are all constructed based on the raw data. However there may exist noise and outlier in the raw data, e.g. face images variations caused by illumination, color and pose. These factors would reduce the ability to represent the correlation among data.

4. Double Nuclear norm based Low Rank model on Grassmann manifold

In this section, we will introduce the formulation and optimization of the proposed G-DNLR model in detail.

4.1. Models Formulation

For a set of Grassmann samples $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$ where $\mathbf{Y}_i \in \mathcal{G}(p, m)$, we could formulate a double nuclear norm based low rank representation model on Grassmann manifold as below:

$$\min_{\mathbf{X},\mathbf{A},\mathbf{B}} \quad \lambda(\|\mathbf{A}\|_* + \|\mathbf{B}\|_*) + \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}}, \quad \text{s.t.} \quad \mathbf{X} = \mathbf{AB}, \tag{14}$$

where $\mathbf{A} \in \mathbb{R}^{n\times r}$, $\mathbf{B} \in \mathbb{R}^{r\times n}$, $r < n$ represents the expect rank of $\mathbf{X}$. $\|\mathbf{A}\|_* + \|\mathbf{B}\|_*$ represents the double nuclear norm for $\mathbf{X}$. We call this model the Double Nuclear norm based Low Rank model on Grassmann manifold (G-DNLR). The objective function is hard to solve owing to the non-linear metric on Grassmann manifold. According to the property and definition in Section 2, we could use the embedding distance to replace the construction error in (14) as below:

$$\begin{aligned}\|\mathbf{Y}_i \ominus \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}\|_{\mathcal{G}} &= \text{dist}_g^2(\mathbf{Y}_i, \biguplus_{j=1}^{n} \mathbf{Y}_j \circledast x_{ji}) \\ &= \|\mathbf{Y}_i\mathbf{Y}_i^T - \sum_{j=1}^{n} x_{ji}\mathbf{Y}_j\mathbf{Y}_j^T\|_F^2.\end{aligned} \tag{15}$$

With this measurement, the function in (14) could be rewritten as below:

$$\min_{\mathbf{X},\mathbf{A},\mathbf{B}} \quad \lambda(\|\mathbf{A}\|_* + \|\mathbf{B}\|_*) + \|\mathbf{Y}_i\mathbf{Y}_i^T - \sum_{j=1}^{n} x_{ji}\mathbf{Y}_j\mathbf{Y}_j^T\|_F^2. \quad \text{s.t.} \quad \mathbf{X} = \mathbf{AB}. \tag{16}$$

Donating $g_{ij} = \text{tr}((\mathbf{Y}_j^T\mathbf{Y}_i)(\mathbf{Y}_i^T\mathbf{Y}_j))$ according to [33], we could rewrite (16) as below:

$$\min_{\mathbf{X},\mathbf{A},\mathbf{B}} \quad \lambda(\|\mathbf{A}\|_* + \|\mathbf{B}\|_*) + \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}), \quad \text{s.t.} \quad \mathbf{X} = \mathbf{AB}, \tag{17}$$

where matrix $\mathbf{G} = \{g_{ij}\}_{n\times n} \in \mathbb{R}^{n\times n}$ is a symmetric matrix. With these transformation, the original non-linear function in (14) could be converted into a linear one.

4.2. Optimization of G-DNLR

G-DNLR model is a complicated optimization problem which is difficult to solve directly. Here, we adopt the alternating direction method of multipliers (ADMM) [3] to solve it. We first introduce two auxiliary variables $\hat{\mathbf{A}} = \mathbf{A}$, $\hat{\mathbf{B}} = \mathbf{B}$ and rewrite (17) as below:

$$\min_{\mathbf{X},\mathbf{A},\mathbf{B},\hat{\mathbf{A}},\hat{\mathbf{B}}} \quad \lambda(\|\hat{\mathbf{A}}\|_* + \|\hat{\mathbf{B}}\|_*) + \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}), \quad \text{s.t.} \quad \mathbf{X} = \mathbf{AB}, \hat{\mathbf{A}} = \mathbf{A}, \hat{\mathbf{B}} = \mathbf{B}. \tag{18}$$

Then we remove the linear equality constraints in (18) by using the augmented Lagrangian method and have

$$\begin{aligned}\min_{\mathbf{X},\mathbf{A},\mathbf{B},\hat{\mathbf{A}},\hat{\mathbf{B}}} \quad & \lambda(\|\hat{\mathbf{A}}\|_* + \|\hat{\mathbf{B}}\|_*) + \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}) \\ & + \text{tr}(\mathbf{F}_1^T(\hat{\mathbf{A}} - \mathbf{A})) + \text{tr}(\mathbf{F}_2^T(\hat{\mathbf{B}} - \mathbf{B})) \\ & + \text{tr}(\mathbf{F}_3^T\mathbf{X} - \mathbf{AB}) + \frac{\gamma}{2}(\|\hat{\mathbf{A}} - \mathbf{A}\|_F^2 \\ & + \|\hat{\mathbf{B}} - \mathbf{B}\|_F^2 + \|\mathbf{X} - \mathbf{AB}\|_F^2),\end{aligned} \tag{19}$$

where $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ are Lagrangian multipliers. $\gamma > 0$ is a penalty parameter. For this problem, $\mathbf{X}, \mathbf{A}, \mathbf{B}, \hat{\mathbf{A}}, \hat{\mathbf{B}}$ and other parameters can be solved by the following alternative iterations, in which superscript t denotes the current iteration step.

4.2.1 Update $\hat{\mathbf{A}}$ with fixing others

When other variables are fixed, (19) degenerates into a function with respect to $\hat{\mathbf{A}}$ as below:

$$\min_{\hat{\mathbf{A}}} \lambda\|\hat{\mathbf{A}}\|_* + \frac{\gamma}{2}\|\hat{\mathbf{A}} - (\mathbf{A} - \frac{\mathbf{F}_1}{\gamma})\|_F^2. \tag{20}$$

We can update $\hat{\mathbf{A}}$ based on the closed-form solution [4]:

$$\hat{\mathbf{A}}^{(t+1)} = \mathbf{U}_1^{(t)}\max\{\mathbf{\Sigma}_1^{(t)} - \frac{\lambda}{\gamma^{(t)}}, 0\}\mathbf{V}_1^{(t)^T}, \tag{21}$$

where $\mathbf{U}_1^{(t)}\mathbf{\Sigma}_1^{(t)}\mathbf{V}_1^{(t)^T}$ is the singular value decomposition (SVD) of $\mathbf{A}^{(t)} - \frac{\mathbf{F}_1^{(t)}}{\gamma^{(t)}}$.

4.2.2 Update $\hat{\mathbf{B}}$ with fixing others

When other variables are fixed, (19) degenerates into a function with respect to $\hat{\mathbf{B}}$ as below:

$$\min_{\hat{\mathbf{B}}} \lambda\|\hat{\mathbf{B}}\|_* + \frac{\gamma}{2}\|\hat{\mathbf{B}} - (\mathbf{B} - \frac{\mathbf{F}_2}{\gamma})\|_F^2. \tag{22}$$

We can update $\hat{\mathbf{B}}$ based on the closed-form solution [4]:

$$\hat{\mathbf{B}}^{(t+1)} = \mathbf{U}_2^{(t)}\max\{\mathbf{\Sigma}_2^{(t)} - \frac{\lambda}{\gamma^{(t)}}, 0\}\mathbf{V}_2^{(t)^T}, \tag{23}$$

where $\mathbf{U}_2^{(t)}\mathbf{\Sigma}_2^{(t)}\mathbf{V}_2^{(t)^T}$ is the singular value decomposition of $\mathbf{B}^{(t)} - \frac{\mathbf{F}_2^{(t)}}{\gamma^{(t)}}$.

4.2.3 Update A with fixing others

When other variables are fixed, (19) degenerates into a function with respect to $\mathbf{A}$ as below:

$$\min_{\mathbf{A}} \|\mathbf{A} - (\hat{\mathbf{A}} + \frac{\mathbf{F}_1}{\gamma})\|_F^2 + \|\mathbf{AB} - (\mathbf{X} + \frac{\mathbf{F}_3}{\gamma})\|_F^2. \quad (24)$$

The closed-form solution to the problem(24) is given by

$$\mathbf{A}^{(t+1)} = (\mathbf{P}_1^{(t)} + \mathbf{P}_2^{(t)}\mathbf{B}^{(t)})(\mathbf{I}_1 + \mathbf{B}^{(t)}\mathbf{B}^{(t)^T})^{-1}, \quad (25)$$

where $\mathbf{P}_1^{(t)} = \hat{\mathbf{A}}^{(t)} + \frac{\mathbf{F}_1^{(t)}}{\gamma^{(t)}}$, $\mathbf{P}_2^{(t)} = \mathbf{X}^{(t)} + \frac{\mathbf{F}_3^{(t)}}{\gamma^{(t)}}$, $\mathbf{I}_1 \in \mathbb{R}^{r\times r}$ represents the identify matrix.

4.2.4 Update B with fixing others

When other variables are fixed, (19) degenerates into a function with respect to $\mathbf{A}$ as below:

$$\min_{\mathbf{B}} \|\mathbf{B} - (\hat{\mathbf{B}} + \frac{\mathbf{F}_2}{\gamma})\|_F^2 + \|\mathbf{AB} - (\mathbf{X} + \frac{\mathbf{F}_3}{\gamma})\|_F^2. \quad (26)$$

The closed-form solution to the problem(26) is given by

$$\mathbf{B}^{(t+1)} = (\mathbf{A}^{(t)^T}\mathbf{A}^{(t)} + \mathbf{I}_1)^{-1}(\mathbf{A}^{(t)^T}\mathbf{P}_2^{(t)} + \mathbf{P}_3^{(t)}), \quad (27)$$

where $\mathbf{P}_3^{(t)} = \hat{\mathbf{B}}^{(t)} + \frac{\mathbf{F}_2^{(t)}}{\gamma^{(t)}}$,

4.2.5 Update X with fixing others

When other variables are fixed, (19) degenerates into a function with respect to $\mathbf{X}$ as below:

$$\min_{\mathbf{X}} \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}) + \frac{\gamma}{2}\|\mathbf{X} - \mathbf{AB} + \frac{\mathbf{F}_3}{\gamma}\|_F^2. \quad (28)$$

The closed-form solution to the problem(28) is given by

$$\mathbf{X}^{(t+1)} = (2\mathbf{G} + \gamma^{(t)}\mathbf{I}_2)^{-1}(2\mathbf{G} + \gamma^{(t)}(\mathbf{A}^{(t)}\mathbf{B}^{(t)} - \frac{\mathbf{F}_3^{(t)}}{\gamma^{(t)}})). \quad (29)$$

4.2.6 Update $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ and γ

The Lagrangian multipliers $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ and penalty parameter γ could be updated as follows:

$$\mathbf{F}_1^{(t+1)} = \mathbf{F}_1^{(t)} + \gamma^{(t)}(\hat{\mathbf{A}}^{(t+1)} - \mathbf{A}^{(t+1)}). \quad (30)$$

$$\mathbf{F}_2^{(t+1)} = \mathbf{F}_2^{(t)} + \gamma^{(t)}(\hat{\mathbf{B}}^{(t+1)} - \mathbf{B}^{(t+1)}). \quad (31)$$

$$\mathbf{F}_3^{(t+1)} = \mathbf{F}_3^{(t)} + \gamma^{(t)}(\mathbf{X}^{(t+1)} - \mathbf{A}^{(t+1)}\mathbf{B}^{(t+1)}). \quad (32)$$

$$\gamma^{(t+1)} = \min(\rho\gamma^{(t)}, \gamma^{\max}), \quad (33)$$

where $\rho > 1$ is a constant and $\gamma^{\max}$ is the upper bound of γ. In our algorithm, the stopping criterion is measured by the following condition:

$$\max\left\{\begin{array}{c} \|\hat{\mathbf{A}}^{(t+1)} - \mathbf{A}^{(t+1)}\|_\infty, \\ \|\hat{\mathbf{B}}^{(t+1)} - \mathbf{B}^{(t+1)}\|_\infty, \\ \|\mathbf{X}^{(t+1)} - \mathbf{A}^{(t+1)}\mathbf{B}^{(t+1)}\|_\infty \end{array}\right\} \le \varepsilon. \quad (34)$$

5. Adaptive Laplacian regularized G-DNLR model

5.1. Model formulation

Combined the Laplacian regularizer, we could construct an affinity matrix $\mathbf{W}$ based on the raw data samples and formulate a Laplacian double nuclear norm based low rank model for Grassmann manifold:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{A},\mathbf{B}} \quad & \lambda(\|\mathbf{A}\|_* + \|\mathbf{B}\|_*) + \alpha\text{tr}(\mathbf{X}\mathbf{L}_{\mathbf{W}}\mathbf{X}^T) \\ & + \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}), \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{AB}. \end{aligned} \quad (35)$$

However, as we discussed in Section 3, the affinity matrix constructed by raw data would be biased by the noise or outliers among the data. Therefore, we adopt the similar approach in [11] to construct an adaptive Laplacian double nuclear norm base low rank model for Grassmann manifold (G-ALDNLR) as follows:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{A},\mathbf{B},\mathbf{W}} \quad & \lambda(\|\mathbf{A}\|_* + \|\mathbf{B}\|_*) + \alpha\text{tr}(\mathbf{X}\mathbf{L}_{\mathbf{W}}\mathbf{X}^T) + \beta\|\mathbf{W}\|_F^2 \\ & + \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}), \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{AB}, \mathbf{W}^T\mathbf{1}^n = \mathbf{1}^n, \mathbf{W} = \mathbf{W}^T, \\ & w_{ij} \ge 0, \forall i,j, \end{aligned} \quad (36)$$

where $\|\mathbf{W}\|_F^2$ is a regularisation term on $\mathbf{W}$ to prevent trivial solution, $\mathbf{1}^n$ represents an n-dimension vector of all 1s. In this formulation, the affinity matrix $\mathbf{W}$ is no longer constructed from the raw data directly. It can be regarded as a latent affinity matrix to reflect the geometry property of the original data space, i.e. the element w_{ij} of $\mathbf{W}$ represents the probability of the i-th and j-th data points belonging to the same class and can be adaptively adjusted in the above optimal procedure.

5.2. Optimization of G-ALDNLR

To solve G-ALDNLR model in (36), we introduce three auxiliary variables $\hat{\mathbf{A}} = \mathbf{A}, \hat{\mathbf{B}} = \mathbf{B}, \mathbf{Z} = \mathbf{X}$ and remove the linear equality constraints rewrite (36) by using the augmented Lagrangian method as below:

$$\begin{aligned} \min_{\Theta} \quad & \lambda(\|\hat{\mathbf{A}}\|_* + \|\hat{\mathbf{B}}\|_*) + \alpha\text{tr}(\mathbf{Z}\mathbf{L}_{\mathbf{W}}\mathbf{Z}^T) + \beta\|\mathbf{W}\|_F^2 \\ & + \text{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\text{tr}(\mathbf{G}\mathbf{X}) + \text{tr}(\mathbf{F}_1^T(\hat{\mathbf{A}} - \mathbf{A})) \\ & + \text{tr}(\mathbf{F}_2^T(\hat{\mathbf{B}} - \mathbf{B})) + \text{tr}(\mathbf{F}_3^T(\mathbf{X} - \mathbf{AB})) \\ & + \text{tr}(\mathbf{F}_4^T(\mathbf{X} - \mathbf{Z})) + \frac{\gamma}{2}(\|\hat{\mathbf{A}} - \mathbf{A}\|_F^2 + \|\hat{\mathbf{B}} - \mathbf{B}\|_F^2 \\ & + \|\mathbf{X} - \mathbf{Z}\|_F^2 + \|\mathbf{X} - \mathbf{AB}\|_F^2), \\ \text{s.t.} \quad & \mathbf{W}^T\mathbf{1}^n = \mathbf{1}^n, \mathbf{W} = \mathbf{W}^T, w_{ij} \ge 0, \forall i,j, \end{aligned} \quad (37)$$

where $\Theta = \{\mathbf{X}, \mathbf{A}, \mathbf{B}, \mathbf{W}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \mathbf{Z}\}$. We could adopt (20) to (27)to solve $\mathbf{A}, \mathbf{B}, \hat{\mathbf{A}}, \hat{\mathbf{B}}$, and solve $\mathbf{Z}, \mathbf{X}, \mathbf{W}$ as below:

5.2.1 Update X with fixing others

When other variables are fixed, (37) degenerates into a function with respect to $\mathbf{X}$ as below:

$$\min_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^T\mathbf{G}\mathbf{X}) - 2\operatorname{tr}(\mathbf{G}\mathbf{X}) + \frac{\gamma}{2}(\|\mathbf{X} - \mathbf{A}\mathbf{B} + \frac{\mathbf{F}_3}{\gamma}\|_F^2 + \|\mathbf{X} - \mathbf{Z} + \frac{\mathbf{F}_4}{\gamma}\|_F^2). \tag{38}$$

The closed-form solution to the problem(38) is given by

$$\mathbf{X}^{(t+1)} = (2\mathbf{G} + \gamma^{(t)}\mathbf{I}_2)^{-1}(2\mathbf{G} + \gamma^{(t)}(\mathbf{P}_4^{(t)} + \mathbf{P}_5^{(t)}), \tag{39}$$

where $\mathbf{P}_4^{(t)} = \mathbf{A}^{(t)}\mathbf{B}^{(t)} - \frac{\mathbf{F}_3^{(t)}}{\gamma^{(t)}}$, $\mathbf{P}_5^{(t)} = \mathbf{X}^{(t)} - \frac{\mathbf{F}_4^{(t)}}{\gamma^{(t)}}$.

5.2.2 Update Z with fixing others

When other variables are fixed, (37) degenerates into a function with respect to $\mathbf{Z}$ as below:

$$\min_{\mathbf{Z}} \ \alpha\operatorname{tr}(\mathbf{Z}\mathbf{L}_{\mathbf{W}}\mathbf{Z}^T) + \frac{\gamma}{2}\|\mathbf{Z} - (\mathbf{X} + \frac{\mathbf{F}_4}{\gamma})\|_F^2. \tag{40}$$

The closed-form solution to the problem(38) is given by:

$$\mathbf{Z}^{(t+1)} = \gamma^{(t)}(\mathbf{Z}^{(t)} + \frac{\mathbf{F}_4^{(t)}}{\gamma^{(t)}})(2\alpha\mathbf{L}_{\mathbf{W}}{}^{(t)} + \gamma^{(t)}\mathbf{I}_2)^{-1}. \tag{41}$$

5.2.3 Update W with fixing others

When other variables are fixed, (37) degenerates into a function with respect to $\mathbf{W}$ as below:

$$\begin{aligned} &\min_{\mathbf{W}} \alpha\operatorname{tr}(\mathbf{Z}\mathbf{L}_{\mathbf{W}}\mathbf{Z}^T) + \beta\|\mathbf{W}\|_F^2, \\ &\text{s.t. } \mathbf{W}^T\mathbf{1}^n = \mathbf{1}^n,\ \mathbf{W} = \mathbf{W}^T,\ w_{ij} \geq 0, \forall i, j. \end{aligned} \tag{42}$$

This problem can be separated into a set of independent subproblems, *i.e.*:

$$\begin{aligned} \mathbf{W}^{(t+1)} = \arg\min_{\mathbf{W}} &\|\mathbf{W} + \mathbf{Q}^{(t)}\|_F^2, \\ &\text{s.t. } \mathbf{W}^T\mathbf{1}^n = \mathbf{1}^n,\ w_{ij} \geq 0, \forall i, j, \end{aligned} \tag{43}$$

where each element $q_{ij}^{(t)}$ of matrix $\mathbf{Q}^{(t)}$ is defined as follows:

$$q_{ij}^{(t)} = \frac{\alpha}{4\beta}\|\mathbf{z}_i^{(t)} - \mathbf{z}_j^{(t)}\|_2^2, \tag{44}$$

where $\mathbf{z}_i^{(t)}$ represents the i-th column of $\mathbf{Z}^{(t)}$. We have the closed-form solution for the i-th column $\mathbf{w}_i^{(t+1)}$ of $\mathbf{W}^{(t+1)}$ by its k nearest neighbours as below:

$$\mathbf{w}_i^{(t+1)} = \left(\frac{1 + \sum_{j=1}^{k}\hat{q}_{ji}^{(t)}}{k}\mathbf{1}^n - \mathbf{q}_i^{(t)}\right)_+, \tag{45}$$

Algorithm 1 The solution to G-DNLR and G-ALDNLR

Require: The Grassmann sample set $\mathcal{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n\}$, the expect rank r, the number of neighbours k, the parameters λ, α, β

Initialize : $\mathbf{X}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{0} \in \mathbb{R}^{n\times n}$, $\mathbf{A}^{(0)} = \hat{\mathbf{A}}^{(0)} = \mathbf{0} \in \mathbb{R}^{n\times r}$, $\mathbf{B}^{(0)} = \hat{\mathbf{B}}^{(0)} = \mathbf{0} \in \mathbb{R}^{r\times n}$, $\mathbf{F}_1^{(0)} = \mathbf{1} \in \mathbb{R}^{n\times r}$, $\mathbf{F}_2^{(0)} = \mathbf{1} \in \mathbb{R}^{r\times n}$, $\mathbf{F}_3^{(0)} = \mathbf{F}_4^{(0)} = \mathbf{1} \in \mathbb{R}^{n\times n}$, initialize $\mathbf{W}$ by solving (49), $\gamma^{(0)} = 10^{-4}$, $\rho > 1$, $\gamma^{max} = 10^{10}$, $\varepsilon = 10^{-7}$, the number of maximum iteration $MaxIter = 1000$

Calculate matrix $\mathbf{G}$ by $g_{ij} = \operatorname{tr}((\mathbf{Y}_j^T\mathbf{Y}_i)(\mathbf{Y}_i^T\mathbf{Y}_j))$;

$t = 0$;

while not converged and $t \leq MaxIter$ **do**

  Update $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \mathbf{A}, \mathbf{B}$ by (20) to (27) respectively;

  Update $\mathbf{X}$ by (28) and (29) for G-DNLR, or update $\mathbf{X}$ by (38) and (39) for G-ALDNLR;

  Update $\mathbf{Z}$ and $\mathbf{W}$ by (40) to (46) respectively for G-ALDNLR ;

  Update $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$ and γ by (30), (31), (32), (47) and (33) respectively;

  Check the convergence condition defined as (34) for G-DNLR, or check the condition defined as (48) for G-ALDNLR;

  $t = t + 1$.

end while

Ensure:

The matrices $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Z}, \mathbf{W}$.

where $\hat{q}_{ji}^{(t)}$ is the j-th element of column vector $\hat{\mathbf{q}}_i^{(t)}$, the elements of $\hat{\mathbf{q}}_i^{(t)}$ are same as those of $\mathbf{q}_i^{(t)}$ in ascending order. For any column $\mathbf{u}$, the operator $(\mathbf{u})_+$ turns negative elements in $\mathbf{u}$ to 0 while keeping the rest. The proof of the closed-form solution to (45) could be found in [23]. Finally, we could adjust $\mathbf{A}^{(t+1)}$ which is generally an unbalanced digraph obtained by (45) as follows:

$$\mathbf{W}^{(t+1)} = \frac{1}{2}(\mathbf{W}^{(t+1)} + \mathbf{W}^{(t+1)^T}). \tag{46}$$

5.2.4 Update $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$, $\mathbf{F}_4$ and γ

$\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ and γ could be updated by (30) to (33). Then we could update $\mathbf{F}_4$ as below:

$$\mathbf{F}_4^{(t+1)} = \mathbf{F}_4^{(t)} + \gamma^{(t)}(\mathbf{X}^{(t+1)} - \mathbf{Z}^{(t+1)}). \tag{47}$$

In this algorithm, the stopping criterion of G-ALDNLR is measured by the following condition:

$$\max\left\{\begin{array}{c} \|\hat{\mathbf{A}}^{(t+1)} - \mathbf{A}^{(t+1)}\|_\infty, \\ \|\hat{\mathbf{B}}^{(t+1)} - \mathbf{B}^{(t+1)}\|_\infty, \\ \|\mathbf{X}^{(t+1)} - \mathbf{Z}^{(t+1)}\|_\infty, \\ \|\mathbf{X}^{(t+1)} - \mathbf{A}^{(t+1)}\mathbf{B}^{(t+1)}\|_\infty \end{array}\right\} \leq \varepsilon. \tag{48}$$

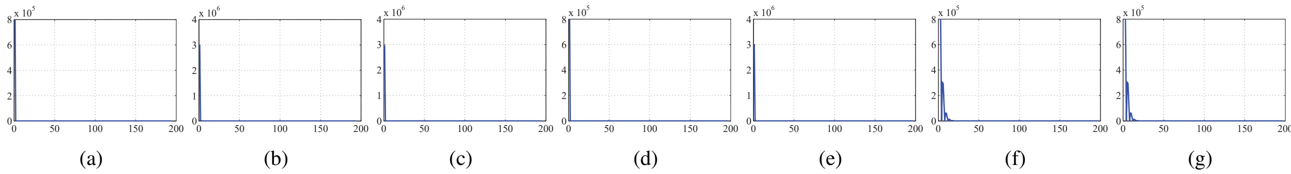


Figure 1. The convergence curves of G-DNLR and G-ALDNLR on Extended Yale B face dataset. In each subfigure, the x-axis represents the iterations and the y-axis represents the value of function: (a) $\|\hat{\mathbf{A}}^{(t+1)} - \mathbf{A}^{(t+1)}\|_\infty$ for G-DNLR; (b) $\|\hat{\mathbf{B}}^{(t+1)} - \mathbf{B}^{(t+1)}\|_\infty$ for G-DNLR; (c) $\|\mathbf{X}^{(t+1)} - \mathbf{A}^{(t+1)}\mathbf{B}^{(t+1)}\|_\infty$ for G-DNLR. (d) $\|\hat{\mathbf{A}}^{(t+1)} - \mathbf{A}^{(t+1)}\|_\infty$ for G-ALDNLR; (e) $\|\hat{\mathbf{B}}^{(t+1)} - \mathbf{B}^{(t+1)}\|_\infty$ for G-ALDNLR; (f) $\|\mathbf{X}^{(t+1)} - \mathbf{Z}^{(t+1)}\|_\infty$ for G-ALDNLR; (g) $\|\mathbf{X}^{(t+1)} - \mathbf{A}^{(t+1)}\mathbf{B}^{(t+1)}\|_\infty$ for G-ALDNLR.

We summarized all update steps in Algorithm 1 for both G-DNLR and G-ALDNLR. For G-ALDNLR, we initialize $\mathbf{W}$ based on the distance of original Grassmann data sample as below:

$$\begin{aligned} \min_{\mathbf{W}} \quad & w_{ij}\|\mathbf{Y}_i\mathbf{Y}_i^T - \mathbf{Y}_j\mathbf{Y}_j^T\|_F^2, \\ \text{s.t.} \quad & \mathbf{W}^T\mathbf{1}^n = \mathbf{1}^n,\ \mathbf{W} = \mathbf{W}^T,\ w_{ij} \geq 0, \forall i, j. \end{aligned} \tag{49}$$

The solution to this problem is similar to (42).

5.3. Converge and Complexity Analysis

For Algorithm 1, by splitting variables in the proposed objective functions, we have a set of subproblems for each variable with closed-form solution. Therefore, all the convergence analysis and proof in [27, 42] could be applicable for our proposed methods. Figure 1 shows an example of the convergence of the proposed methods on Extended Yale B face dataset[14]. It is indicated that the curves decrease fast and almost tend to be stable within 200 iterations, which verifies the good convergence of our algorithm.

Further, we discuss the complexity of the proposed models. Calculating $\mathbf{G}$ has a complexity of $\mathcal{O}(n^2(p^2m + p^3))$. In each iteration step, the complexity of updating $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\mathbf{A}$ and $\mathbf{B}$ are all $\mathcal{O}(nr^2)$. The complexity of updating $\mathbf{Z}$ and $\mathbf{X}$ are both $\mathcal{O}(n^3)$. The complexity of updating $\mathbf{W}$ is $\mathcal{O}(n^2)$. Therefore, the total complexity of G-DNLR and G-ALDNLR are $\mathcal{O}(n^2(p^2m + p^3) + t(4nr^2 + n^3))$ and $\mathcal{O}(n^2(p^2m + p^3) + t(4nr^2 + 2n^3 + n^2))$ respectively. We also list the complexities of other methods in Table 1. Besides, we test all the methods on Extended Yale B dataset and show the wall-clock time in Table 1. It demonstrates that our proposed methods (G-DNLR & G-ALDNLR) have acceptable executive time. All methods are coded in Matlab R2014a and implemented on an Intel Core i7-7700 3.60GHz CPU machine with 16G RAM.

Method	Complextiy	Running Time
SSC	$\mathcal{O}(tn^2(1+n))$	1.61
LRR	$\mathcal{O}(2tn^3)$	21.41
LS3C	$\mathcal{O}(tn^2(sn^2+1))$	30.65
SCGSM	$\mathcal{O}(m^3p^3d^2t(n+d))$	1050.67
G-KM	$\mathcal{O}(n^2p^2(m+p)+3n^3$	0.11
G-CLRSR	$\mathcal{O}(n^2p^2(m+p)+tn^2(3n+1))$	90.56
G-LRR	$\mathcal{O}(n^2p^2(m+p)+2tn^3)$	17.76
G-LLRR	$\mathcal{O}(n^2p^2(m+p)+2tn^3)$	18.46
G-PSSVLR	$\mathcal{O}(n^2p^2(m+p)+2tn^3)$	15.17
G-LPSSVLR	$\mathcal{O}(n^2p^2(m+p)+2tn^3)$	15.74
G-DNLR	$\mathcal{O}(n^2p^2(m+p)+tn(4r^2+n^2))$	17.28
G-ALDNLR	$\mathcal{O}(n^2p^2(m+p)+tn(4r^2+2n^2+n))$	32.25

Table 1. The complexity and running time (second) on Extended Yale B dataset of various methods.

6. Empirical Comparison

We test our methods on four datasets, including the Extended Yale B face dataset[14], CMU-PIE face dataset[30], Ballet action dataset [9] and SKIG gesture dataset [18]. The performance of the proposed methods is compared with some state-of-the-art clustering algorithms, such as SSC [7], LRR [16], LS3C [26], SCGSM [31], G-KM [29], G-LRR [33], G-LLRR [34, 36], G-PSSVLR &G-LPSSVLR[37] and G-CLRSR [35]. In our methods, after learning the representation $\mathbf{X}$, we use the NCut method [28] to obtain the final clustering results. The clustering results are measured by the clustering Accuracy (ACC) and Normalized Mutual Information (NMI). The details of data setting and results analysis are given below.

6.1. Data and parameters setting

The Extended Yale B dataset contains 2,414 frontal face images of 38 subjects. Each subject has about 64 images. In our experiments, we resize images into 20×20 and randomly select 8 images from the same subject to form an imageset sample. The CMU-PIE dataset is composed of 68 subjects and each subject has 21 front face images with no expression. Every 4 images from the same subject are selected to form an imageset sample. Ballet action contains 44 video clips collected from an instructional Ballet DVD. We resize each image into 30×30 and every 12 images are chosen for an imageset from each clip. SKIG gesture dataset consists of 1080 RGB-D video collected from 6 subjects with 10 gesture types. Each image is resized as 24×32 and each clip is regarded as an imageset.

In our experiments, we first transform each image into a m-dimension vector. For vector based LRR, SSC and LS3C methods, we stack all image vectors from the same imageset as a long vector and adopt PCA to reduce the dimension which equals to the dimension of PCA components retaining 95% of its variance energy. For other Grassmann man-

Method	SSC	LRR	LS3C	SCGSM	G-KM	G-CLRSR	G-LRR	G-LLRR	G-PSSVLR	G-LPSSVLR	G-DNLR	G-ALDNLR
Extended Yale B	0.4032	0.4659	0.2461	0.7946	0.8365	0.8194	0.8135	0.7921	0.9035	0.9118	*0.9749*	**0.9855**
CMU-PIE	0.5231	0.4034	0.2761	0.5732	0.6025	0.6289	0.6153	0.5862	0.6213	0.6373	*0.6618*	**0.7244**
Ballet	0.2962	0.2923	0.4262	0.5613	0.5699	0.5931	0.5912	0.6076	0.6013	0.6243	*0.6768*	**0.6885**
SKIG	0.3892	0.2537	0.2941	0.3716	0.5308	0.5083	0.5022	0.5176	0.5502	0.5712	*0.6244*	**0.6667**

Table 2. The accuracy results of various methods on four datasets.

Method	SSC	LRR	LS3C	SCGSM	G-KM	G-CLRSR	G-LRR	G-LLRR	G-PSSVLR	G-LPSSVLR	G-DNLR	G-ALDNLR
Extended Yale B	0.6231	0.6813	0.4992	0.9326	0.9341	0.9103	0.8903	0.8923	0.9262	0.9461	*0.9874*	**0.9891**
CMU-PIE	0.7865	0.7321	0.6313	0.5736	0.5976	0.8132	0.8103	0.7942	0.7926	0.8080	*0.8510*	**0.8774**
Ballet	0.2813	0.2910	0.4370	0.5646	0.5779	0.5862	0.5762	0.5983	0.5837	0.6102	*0.6310*	**0.6692**
SKIG	0.4762	0.3343	0.3421	0.5367	0.5671	0.5679	0.5450	0.5571	0.5692	0.5873	*0.6482*	**0.6484**

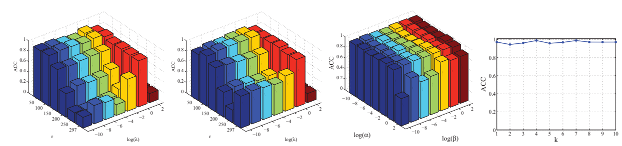
Table 3. The normalized mutual information results of various methods on four datasets.

ifold based methods, we form all image vectors from the same imageset as a matrix. Then SVD is applied on the matrix and we pick up the first p columns of the left singular matrix as a sample data on Grassmann manifold $\mathcal{G}(p, m)$.

To get a suitable setting for these parameters, we study each parameter's influence on the clustering accuracy by some pre-experiments. In these experiments, each parameter is tested with fixing the other parameters for both G-DNLR and G-ALDNLR. We tune the rank value r within $\{1, 2, \cdots, n\}$, the neighbour number k within $\{1, 2, \cdots, 10\}$, other parameters λ, α, β are tuned within $\{10^{-10}, 10^{-9}, \cdots, 1, 10, 10^2\}$. Figure 2 shows the influence of $\lambda, \alpha, \beta, r$ and k on Extended Yale B dataset as an example. The parameters setting are: $\lambda = 1$, $r = 100$ for G-DNLR and $\lambda = \alpha = \beta = 10^{-2}, r = 100, k = 7$ for G-ALDNLR on Extended Yale B dataset; $\lambda = 1$, $r = 142$ for G-DNLR and $\lambda = \alpha = \beta = 10^{-2}, r = 110, k = 6$ for G-ALDNLR on CMU-PIE dataset; $\lambda = 10^2, r = 6$ for G-DNLR and $\lambda = \alpha = \beta = 10^{-7}, r = 10, k = 8$ for G-ALDNLR on Ballet dataset; $\lambda = 10^{-2}, r = 27$ for G-DNLR and $\lambda = 1, \alpha = 10^2, \beta = 10^{-5}, r = 27, k = 8$ for G-ALDNLR on SKIG dataset.

6.2. Results analysis

All clustering results are shown in Tables 2 and 3. For each experiment, the clustering are repeated 20 times and the average results are reported. The best results are bold, the second ones are in italic script and the third ones are underlined. From the results, Grassmann manifold representation based methods always have better performances than the vectors based ones (SSC, LRR, LS3C), which is explained that the manifold representation have the advantage of revealing the complicated relationship within the imageset data effectively. In all the methods, the low rank based models always obtain the top results, which shows the benefit of low rank representation. From the results, our proposed G-DNLR and G-ALDNLR obtain the top 2 best results and outperform the third best ones with about 4 and 8 percentage points gap in terms of ACC on avenges respectively. The significant improvement of our method is analyzed and own to the superiority that the proposed method



(a) (b) (c) (d)

Figure 2. The clustering accuracy of the proposed methods on Extended Yale B dataset with different parameters setting: (a) G-DNLR with different λ and r; (b) G-ALDNLR with different λ and r; (c) G-ALDNLR with different α and β; (d) G-ALDNLR with different k.

not only adopts the double nuclear norm but also constructs adaptive affinity matrix based on the local structure.

7. Conclusion

In this paper, we propose two new low rank model on Grassmann manifold for high-dimension data clustering task. Instead of the traditional single nuclear norm, we adopt a kind of Schatten-p quasi-norm named Double Nuclear norm to formulate novel clustering models on Grassmann manifold with non-linear metric, which is called G-DNLR. Further, to exploit the local geometrical structure of the data samples, we integrated the adaptive Laplacian regularization with G-DNLR as G-ALDNLR. The proposed models has been evaluated on four public datasets. The experimental results show that our proposed models outperforms state-of-the-art ones.

Acknowledgement

The research project is supported by National Natural Science Foundation of China under Grant No. 61632006, 61672071, U1811463, 61772048, 61806014, Beijing Natural Science Foundation No. 4172003, 4184082, the Innovation Foundation of Science and Technology of Dalian No. 2018J11CY010.

References

[1] Pierre-Antoine Absil, Robert E. Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[2] Arijit Biswas and David Jacobs. Active image clustering: Seeking constraints from humans to complement algorithms. In *Procceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2152–2159, 2012.

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[4] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2008.

[5] Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5–6):877–905, 2008.

[6] Fan R. K. Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., Providence, RI, USA, 1997.

[7] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):2765–2781, 2013.

[8] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

[9] Alireza Fathi and Greg Mori. Action recognition by learning mid-level motion features. In *Procceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[10] Amit Gruber and Yair Weiss. Multibody factorization with uncertainty and missing data using the em algorithm. In *Procceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 707–714, 2004.

[11] Xiaojie Guo. Robust subspace segmentation by simultaneously learning data representations and their affinity matrix. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 3547–3553, 2015.

[12] Mehrtash Harandi, Conrad Sanderson, Chunhua Shen, and Brian C. Lovell. Dictionary learning and sparse coding on grassmann manifolds: An extrinsic solution. In *Procceedings of International Conference on Computer Vision*, pages 3120–3127, 2013.

[13] Stelios Krinidis and Vassilios Chatzis. A robust fuzzy local information c-means clustering algorithm. *IEEE Transactions on Image Processing*, 19(5):1328–1337, 2010.

[14] Kuang-Chih Lee, Jeffrey Ho, and David J Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[15] Gilad Lerman and Teng Zhang. Robust recovery of multiple subspaces by geometric l_p minimization. *The Annals of Statistics*, 39(5):2686–2715, 2011.

[16] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *Proceedings of International Conference on Machine Learning*, pages 663–670, 2010.

[17] Hongfu Liu, Ming Shao, Sheng Li, and Yun Fu. Infinite ensemble for image clustering. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1745–1754, 2016.

[18] Li Liu and Ling Shao. Learning discriminative representations from RGB-D video data. In *Procceedings of International Joint Conference on Artificial Intelligence*, pages 1493–1500, 2013.

[19] Xiaoqiang Lu, Yulong Wang, and Yuan Yuan. Graph-regularized low-rank representation for destriping of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):4009–4018, 2013.

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

[21] Mohammadreza Malek-Mohammadi, Massoud Babaie-Zadeh, and Mikael Skoglund. Performance guarantees for schatten-p quasi-norm minimization in recovery of low-rank matrices. *Signal Processing*, 114:225–230, 2015.

[22] Feiping Nie, Hua Wang, Heng Huang, and Chris Ding. Joint schatten p-norm and ℓ_p-norm robust matrix completion for missing value recovery. *Knowledge and Information Systems*, 42(3):525–544, 2015.

[23] Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 977–986, 2014.

[24] Feiping Nie, Xiaoqian Wang, Michael I. Jordan, and Heng Huang. The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of AAAI Conference on Artificial Intelligence*, pages 1969–1976, 2016.

[25] Tae-Hyun Oh, Yu-Wing Tai, Jean-Charles Bazin, Hyeongwoo Kim, and In So Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):744–758, 2015.

[26] Vishal M. Patel, Hien Van Nguyen, and René Vidal. Latent space sparse subspace clustering. In *Procceedings of IEEE International Conference on Computer Vision*, pages 225–232, 2013.

[27] Fanhua Shang, James Cheng, Yuanyuan Liu, Zhi-Quan Luo, and Zhouchen Lin. Bilinear factor matrix norm minimization for robust pca: Algorithms and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2066–2080, 2018.

[28] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):888–905, 2010.

[29] Sareh Shirazi, Mehrtash T. Harandi, Conrad Sanderson, Azadeh Alavi, and Brian C. Lovell. Clustering on grassmann manifolds via kernel embedding with application to action analysis. In *Proceedings of IEEE International Conference on Image Processing*, pages 781–784, 2012.

[30] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[31] Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa. Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2273–2286, 2011.

[32] René Vidal. A tutorial on subspace clustering. *IEEE Signal Processing Magazine*, 25(2):52–68, 2011.

[33] Boyue Wang, Yongli Hu, Junbin Gao, Yanfeng Sun, and Baocai Yin. Low rank representation on Grassmann manifolds. In *Proceedings of Asian Conference on Computer Vision*, pages 81–96, 2014.

[34] Boyue Wang, Yongli Hu, Junbin Gao, Yanfeng Sun, and Baocai Yin. Laplacian lrr on product grassmann manifolds for human activity clustering in multicamera video surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3):554–566, 2017.

[35] Boyue Wang, Yongli Hu, Junbin Gao, Yanfeng Sun, and Baocai Yin. Cascaded low rank and sparse representation on grassmann manifolds. In *Proceedings of International Joint Conference on Artificial Intellignce*, pages 2755–2761, 2018.

[36] Boyue Wang, Yongli Hu, Junbin Gao, Yanfeng Sun, and Baocai Yin. Localized lrr on grassmann manifold: An extrinsic view. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10):2524–2536, 2018.

[37] Boyue Wang, Yongli Hu, Junbin Gao, Yanfeng Sun, and Baocai Yin. Partial sum minimization of singular values representation on grassmann manifolds. *ACM Transactions on Knowledge Discovery from Data*, 12(1):Article 13, 2018.

[38] Shusen Wang, Dehua Liu, and Zhihua Zhang. Nonconvex relaxation approaches to robust matrix recovery. In *Procceedings of International Joint Conference on Artificial Intelligence*, pages 1764–1770, 2013.

[39] Tong Wu and Waheed U. Bajwa. Learning the nonlinear geometry of high-dimensional data: Models and algorithms. *IEEE Transactions on Signal Processing*, 63(23):6229–6244, 2015.

[40] Ming Yin, Junbin Gao, and Zhouchen Lin. Laplacian regularized low-rank representation and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):504–517, 2016.

[41] Chong You, Daniel Robinson, and René Vidal. Scalable sparse subspace clustering by orthogonal matching pursuit. In *Procceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3918–3927, 2016.

[42] Hengmin Zhang, Jian Yang, Fanhua Shang, Chen Gong, and Zhenyu Zhang. LRR for subspace segmentation via a tractable schatten-p norm minimization and factorization. *IEEE Transactions on Cybernetics*, 49(5):172–1734, 2019.