Geometric Understanding for Unsupervised Subspace Learning

Shihui Ying[1] , **Lipeng Cai**[1] , **Changzhou He**[2] and **Yaxin Peng**[1*]
[1]Department of Mathematics, School of Science, Shanghai University, Shanghai, China
[2]Qualcomm (Shanghai) Co. Ltd., China
{shying, xiaocaibao77}@shu.edu.cn, changzhouhe@163.com, yaxin.peng@shu.edu.cn

Abstract

In this paper, we address the unsupervised subspace learning from a geometric viewpoint. First, we formulate the subspace learning as an inverse problem on Grassmannian manifold by considering all subspaces as points on it. Then, to make the model computable, we parameterize the Grassmannian manifold by using an orbit of rotation group action on all standard subspaces, which are spanned by the orthonormal basis. Further, to improve the robustness, we introduce a low-rank regularizer which makes the dimension of subspace as low as possible. Thus, the subspace learning problem is transferred to a minimization problem with variables of rotation and dimension. Then, we adopt the alternately iterative strategy to optimize the variables, where a structure-preserving method, based on the geodesic structure of the rotation group, is designed to update the rotation. Finally, we compare the proposed approach with six state-of-the-art methods on two different kinds of real datasets. The experimental results validate that our proposed method outperforms all compared methods.

1 Introduction

Subspace learning is a kind of highly effective approaches to dimensionality reduction, which is widely applied in multi-view data analysis [Ding and Fu, 2018], image classification [Fang *et al.*, 2018], feature representation [Li *et al.*, 2015], etc. In recent years, it becomes one of the mutual topics in computer vision, signal processing and statistical learning.

The goal of subspace learning is to map the high-dimensional data to a low-dimensional space for representing more robust features while retaining as much information as possible. The basic idea is directly mining a latent low-dimensional manifold from the data, such that the local topological and/or geometric structures are preserved as much as possible. From the availability of training samples, subspace learning can be traditionally divided into three categories: supervised, unsupervised, and semi-supervised cases. Due to the fundamentality and simplification, in recent years, unsupervised subspace learning attracts more and more attention [Vaswani *et al.*, 2018].

As a classical unsupervised approach, principal component analysis (PCA) finds the principal components, which are some orthogonal basis vectors most of data lies on, and all principal components span a subspace that well represents the data [Turk and Pentland, 1991]. Although PCA well concerns the globally linear structure of data, it does not catch its latently geometric structure. On the other hand, it is sensitive to the noise and outliers.

Considering the local geometric structure of data, Tenenbaum et al. [2000], Saul and Roweis [2003], and Belkin and Niyogi [2003] propose Isometric Mapping (ISOMAP), Locally Linear Embedding (LLE), and Laplacian Eigenmap (LE), respectively, by introducing the manifold assumption for data distribution. All these three approaches can effectively discover the locally geometric structure of the data by preserving data similarity before and after mapping, but they only provide the embedding results of training samples. Therefore, they cannot be extended to the classification problems. Moreover, they have high computational cost. For this, He et al. [2005b] establish the Locality Preserving Projection (LPP) method by approximating LE method, which well balances the manifold and the local structure of the data. Also considering the local structure, they propose a linear method, named Neighborhood Preserving Embedding (NPE), and apply it in classification problem [He *et al.*, 2005a]. Later, Qiao et al. [2013] develop an explicit nonlinear mapping for manifold learning. More recently, Zhu et al. [2018] design an unsupervised spectral feature selection method by the self-expressiveness of the features for preserving the local structure of features, and a low-rank constraint on the weight matrix to preserve the global structure.

On the other hand, to improve the robustness of PCA, in recent years, sparse representation (SR) [Wright *et al.*, 2009] and low-rank representation (LRR) [Liu *et al.*, 2013; Li and Fu, 2016] based methods are established. The SR-based methods can obtain a good robustness for the noisy data by seeking a best sparse representation for each group of samples, but they omit the global property of the data. For example, with the assumption of one subspace distribution, Candès et al. [2011] propose the Robust PCA (RPCA) method, which decomposes the data into low-rank background and s-

*Contact Author

parse noise parts, and hence greatly promotes the robustness of data recovery. But in amount of real circumstances, data are not located in one subspace. Therefore, subspace clustering problem is proposed to describe the multi-subspace case. Elhamifar and Vidal [2013] discuss this task. Therein, Liu et al. [2010] adopt the nuclear norm for the wight matrix, which makes the acquirement of global representation easier. Further, to fit for data-missing case, Liu and Yan [2011] develop the Latent LRR (LatLRR) method. Later, Li and Fu [2016] develop a supervised regularization-based robust subspace (SRRS) method based on the label information and low-rank representation. More recently, Ding and Fu [2018] improve it to multi-view data analysis by collective low-rank subspace learning.

Although these methods improve more or less the robustness of subspace learning, they do not fully consider the geometric structure of the set of subspaces, which may further improve the performance of subspace learning. For example, Hamm and Lee [2008] propose a unifying view on the subspace learning method by formulating the problems as an optimization problem on the Grassmannian manifold, and performs feature extraction and classification in the same space. Li et al. [2008] develop an incremental subspace learning method by using the Log-Euclidean metric. Later, by using the non-Euclidean framework, Huang et al. [2014] consider the classification problem. More recently, Hauberg et al. [2014; 2016] find that the Karcher average of all one-dimensional subspaces spanned by normally distributed data coincides with the first principal component. Further, Chakraborty et al. [2017] extend the Karcher average computation for all one-dimensional subspaces to K-dimensional case. All these works validate the efficiency of Grassmannian manifold based methods, but there are still two shortages as follows. 1) All these methods need to have the dimension of subspaces as a prior knowledge, or need to enumerate all possible dimensions for subspace learning. 2) Grassmannian manifold is a graded manifold, hence it is difficult to represent the subspaces of different dimensions in the same framework.

Therefore, in this paper we will transfer the subspace learning on the Grassmannian manifold to the minimization problem on the rotation group, by representing all subspaces with different dimensions as an orbit of rotation group action on some standard subspaces. These standard subspaces are spanned by the orthonormal basis. Further, to avoid enumerating all possible dimensions for subspace learning, we introduce a low-rank regularizer which makes the dimension of subspace as low as possible.

The rest of this paper is organized as follows. In Section 2, we develop a new model for subspace learning from the rotation group action viewpoint. Then, an alternately iterative strategy and a structure-preserving algorithm are designed in Section 3. In Section 4, we demonstrate the effectiveness of our proposed method, and compare it with six state-of-the-art methods on two different kinds of real datasets (COIL-100, and MNIST). Finally, this paper is concluded in Section 5.

2 Geometric Model for Subspace Learning

We first consider the geometric structure of the rotation group $\mathrm{SO}(d)$. By the definition,

$$\mathrm{SO}(d) = \{R \in \mathbb{R}^{d\times d} | RR^T = R^T R = I_d, \det(R) = 1\},$$

where I_d is the identity matrix of the order d, and $\mathrm{SO}(d)$ is an $m := \frac{d(d-1)}{2}$ dimensional matrix Lie group with its Lie algebra $\mathfrak{so}(d)$ defined by

$$\mathfrak{so}(d) = \{\Omega \in \mathbb{R}^{d\times d} | \Omega^T = -\Omega\},$$

which is the set of all skew-symmetric matrices. The Lie algebra $\mathfrak{so}(d)$ is the linearization of Lie group $\mathrm{SO}(d)$ at the identity.

The goal of subspace learning is to find a latent low-dimensional manifold from the data for representing more robust features, while retaining as much information as possible. Therefore, it is natural to reduce the subspace learning to an optimization problem on the set of all subspaces. On the other hand, from the geometric viewpoint, all subspaces with dimension r of Euclidean spaces $\mathbb{R}^d$ form the Grassmannian manifold

$$\mathrm{Gr}(d, r) = \{\mathbb{S} \subset \mathbb{R}^d | \dim(\mathbb{S}) = r\},$$

where $\mathbb{S}$ is a subspace with the dimension r of $\mathbb{R}^d$, and r is ergodic from 0 to d. Note that the trivial subspaces ($r = 0$ or $r = d$) should be eliminated in subspace learning, and hence we always assume that $1 \le r \le d-1$.

As mentioned above, the Grassmannian manifold $\mathrm{Gr}(d, r)$ is a graded manifold with respect to the dimension r, therefore, the first issue is how to uniformly parameterize this manifold for different r. After that, then, we can address the subspace learning on such manifold. Below, we will give a uniform representation for all subspaces first from the viewpoint of the rotation group action, and then we model the subspace learning as an optimization model on $\mathrm{SO}(d) \ltimes \mathbb{N}$.

2.1 Representation of Subspaces

Figure 1 gives an illustration for subspace learning. Given the set of centered samples $\{x_i\}_{i=1}^n$, we wish to find the best subspace (the hyperplane crossing the origin) such that it best describes the distribution of all samples.

To uniformly represent the subspaces, we adopt the approach of the rotation group $\mathrm{SO}(d)$ action. It is from the fact that there exists a rotation $R \in \mathrm{SO}(d)$, by which any subspace $\mathbb{S}$ with dimension r can be rotated from the standard subspace $\mathbb{S}_0$ spanned by r orthonormal basis. In detail, it is described as follows.

Let $e_i = (0, \cdots, 1, \cdots, 0)^T$ be the i^{th} orthonormal basis, $B_r = \mathrm{span}\{e_1, e_2, \cdots, e_r\}$ be the standard r-dimensional subspace, and $P_r = [e_1, \cdots, e_r]^T$ be the projection from $\mathbb{R}^d$ to B_r. Then, for any r-dimensional subspace $\mathbb{S} \in \mathrm{Gr}(d, r)$, we can find a rotation $R \in \mathrm{SO}(d)$, such that

$$\mathbb{S} = RB_r.$$

That is,

$$\mathrm{Gr}(d, r) = \{RB_r | R \in \mathrm{SO}(d)\}.$$

Then, the unit normal vector N_r of the subspace $\mathbb{S}$ is the rotated unit normal vector of the standard subspace $\mathbb{S}_0$ by R. That is,

$$N_r = RN_{0r},$$

where $N_{0r} = \frac{1}{\sqrt{d-r}}(\underbrace{0, \cdots, 0}_{r}, \underbrace{1, \cdots, 1}_{d-r})^T$.

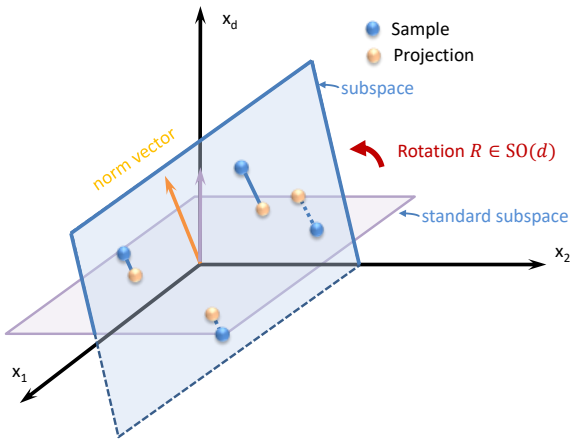


Figure 1: Illustration for subspace learning.

2.2 Model for Subspace Learning

By the representation of rotation group action, we reformulate the subspace learning as follows. As mentioned above, the goal of subspace learning is finding the best subspace such that it best describes the distribution of all samples. Therefore, first we should make the subspace best fitting the samples. That is, we should assume that the total distance of all samples from the subspace is as small as possible, where the distance between samples and the subspace can be formulated by

$$d(x_i, RB_r) = \|x_i^T R N_{0r}\|,$$

where R is the best rotation, B_r is the standard subspace and N_{0r} is the unit normal vector of B_r.

Then, when the dimension of subspace is given by r, one part of model is to minimize the totally squared distance of all samples from the subspace with respect to the rotation R. That is,

$$R^* = \arg\min_{R\in \mathrm{SO}(d)} \sum_{i=1}^{n} \|x_i^T R N_{0r}\|^2. \tag{1}$$

On the other hand, due to the dimensional reduction, we want the dimension of the subspace is sufficiently small. Therefore, we introduce the regularization term to (1), and the model is reformulated by

$$(R^*, r^*) = \arg\min_{R\in \mathrm{SO}(d), 1\le r\le d-1} \sum_{i=1}^{n} \|x_i^T R N_{0r}\|^2 + \lambda r. \tag{2}$$

where λ is a balanced parameter. Therefore, the subspace learning problem is transferred to the optimization problem on the $\mathrm{SO}(d) \ltimes \mathbb{N}$. It is remarkable that the regularity term is a low-rank constraint in previous literature because of $\mathrm{rank}(P_r) = r$. On the other hand, from the Eq. (2), the first term ensures that the samples are as close as possible to the subspace. We know that this error decreases with the increment of r. Therefore, the model is well balanced the error and the dimension.

3 Algorithm

In this section, we will provide the solver for the model (2).

3.1 Iterative Strategy

It is seen that model (2) is an optimization problem with two kinds of independent variables. Therefore, we use an alternately iterative strategy. That is, we decompose the model (2) into two following subproblems.

(S1) For current r^k, we update the next rotation R^{k+1} by

$$R^{k+1} = \arg\min_{R\in \mathrm{SO}(d)} \sum_{i=1}^{n} \|x_i^T R N_{0r^k}\|^2. \tag{3}$$

(S2) For updated rotation R^{k+1}, we update the next dimension r^{k+1} by

$$r^{k+1} = \arg\min_{1\le r\le d-1} \sum_{i=1}^{n} \|x_i^T R^{k+1} N_{0r}\|^2 + \lambda r. \tag{4}$$

Then, by alternately updating Eqs. (3) and (4), we obtain the best rotation R^* and the best dimension r^* of the subspace, and moreover the best subspace $\mathbb{S}^* = R^* P_{r^*}$.

Eq. (4) can be solved by traditional methods, since it is a finite and discrete minimization problem. Then, we will focus on the solution of the subproblem (S1) by the geometric structure of the rotation group $\mathrm{SO}(d)$.

3.2 Structure-Preserving Algorithm

To solve the subproblem (S1), we first consider the exponential map exp from $\mathfrak{so}(d)$ to $\mathrm{SO}(d)$, by which we define the intrinsically iterative format as

$$R^{k+1} = R^k \exp\left(\sum_{j=1}^{m} a_j^k E_j\right), \tag{5}$$

where $\{E_j\}_{1\le j\le m}$ is the basis of Lie algebra $\mathfrak{so}(d)$, and a_j^k is the coefficient at step k. By this format, we insure that R at each step is an exact rotation, and hence it is a structure-preserving iterative method.

Then, we apply this method to solve the subspace (S1). Substituting the iterative format to Eq. (3), we have

$$a^k = \arg\min_{a\in\mathbb{R}^m} \sum_{i=1}^{n} \left\| x_i^T R^k \exp\left(\sum_{j=1}^{m} a_j E_j\right) N_{0r^k} \right\|^2. \tag{6}$$

Therefore, the subproblem (S1) is transferred to the conventional optimization problem. Then, to further simplify the route of solution, we adopt the quadratic approximation for the objective function by using the linearization for the exponential map [Helgason, 1978; Warner, 2013].

$$\exp\left(\sum_{j=1}^{m} a_j E_j\right) \approx Id + \sum_{j=1}^{m} a_j E_j,$$

when a_j is sufficiently small.

Let $G(a)$ be this approximated function, then we have

$$\begin{aligned} G(a) &= \sum_{i=1}^{n} \left\| x_i^T R^k \left(Id + U(a)\right) N_{0r^k} \right\|^2 \\ &= \mathrm{tr}\left(R^k Y^k R^{k^T} X X^T\right) \\ &\quad + \mathrm{tr}\left(R^k Y^k U(a)^T R^{k^T} X X^T\right) \\ &\quad + \mathrm{tr}\left(R^k U(a) Y^k R^{k^T} X X^T\right) \\ &\quad + \mathrm{tr}\left(R^k U(a) Y^k U(a)^T R^{k^T} X X^T\right), \end{aligned} \quad (7)$$

where $U(a) = \sum_{j=1}^{m} a_j E_j$, $Y^k = N_{0r^k} N_{0r^k}^T$, and X is the data matrix $[x_1, x_2, \cdots, x_n]$.

Further, let $Z^k = R^{k^T} X X^T R^k$, then

$$\begin{aligned} G(a) &= \mathrm{tr}\left(Y^k Z^k\right) + \mathrm{tr}\left(Y^k U(a)^T Z^k\right) \\ &\quad + \mathrm{tr}\left(U(a) Y^k Z^k\right) + \mathrm{tr}\left(U(a) Y^k U(a)^T Z^k\right) \\ &= \mathrm{tr}\left(Y^k Z^k\right) + 2\mathrm{tr}\left(U(a) Y^k Z^k\right) \\ &\quad + \mathrm{tr}\left(U(a) Y^k U(a)^T Z^k\right), \end{aligned} \quad (8)$$

where the last equation holds because

$$\begin{aligned} \mathrm{tr}\left(Y^k U(a)^T Z^k\right) &= \mathrm{tr}\left(Z^{k^T} U(a) Y^{k^T}\right) \\ &= \mathrm{tr}\left(U(a) Y^k Z^k\right). \end{aligned}$$

Let $M_{ij}^k = \mathrm{tr}(E_i Y^k E_j^T Z^k)$ and $b_i^k = \mathrm{tr}(Y^k E_i^T Z^k)$, then

$$G(a) = a^T M^k a + 2 b^{k^T} a + \mathrm{tr}\left(Y^k Z^k\right), \quad (9)$$

where M^k is the matrix with element M_{ij}^k, and b^k is the vector with element b_i^k.

By the KKT condition, the optimal coefficients a^k should have the equation $\nabla G(a) = 0$ held. That is,

$$M^k a = -b^k, \quad (10)$$

Therefore, we have

$$a^k = -[M^k]^\dagger b^k. \quad (11)$$

where $[M^k]^\dagger$ is the Moore-Penrose pseudo inverse of M^k. By this approximation, we deduce a closed form for the almost optimal coefficients. It highly improves the computational efficiency.

Therefore, we obtain the algorithm for subspace learning and summarize it as follows.

Algorithm 1 Intrinsic algorithm for subspace learning

Require: Centered dataset $X = \{x_i\}_{i=1}^n$.
Ensure: The best subspace for data representation.
1: Initialize: The orthonormal basis $\{e_i\}$, the basis of Lie algebra $\mathfrak{so}(d)$, maximal iteration T, the balanced parameter λ, $r^0 = d$ and $R^0 = Id$.
2: **for** each $k \in [1, T]$ **do**
3: For current r^k, update the next rotation R^{k+1} by (5), where the coefficient vector a^k is updated by (11).
4: For updated rotation R^{k+1}, update the next dimension r^{k+1} by solving the finite discrete minimization problem (4).
5: **end for**
6: Output: The best subspace $\mathbb{S}^* = R^* B_{r^*}$.

4 Experiment Results

To demonstrate the effectiveness of the proposed subspace learning algorithm, in this section, experiments of classification by using the k-NN classifier on learned subspace (reduced feature space) are conducted on two real datasets, i.e., COIL-100 object dataset [Nene *et al.*, 1996], and MNIST digit dataset [Lecun *et al.*, 1998]. Here, we more focus on the learned subspace, and hence classification results are used to estimate the effects of the selected features. Therefore, for the classification task, we select the k-NN classifier which is more dependent with features. Further, we compare our proposed method with six state-of-the-art subspace learning methods, which include PCA (as a baseline experiment) [Turk and Pentland, 1991], LDA [Belhumeur *et al.*, 1997], LPP [He *et al.*, 2005b], NPE [He *et al.*, 2005a], LSDA [Cai *et al.*, 2007], and SRRS [Li and Fu, 2016]. All programs are written in Matlab 2013a and run by PC with Intel(R) Core(TM) i7-7500U CPU and 32 GB RAM.

4.1 COIL-100 Object Dataset

This dataset contains 100 objects. All images of one object are taken 5 degrees apart because the object is imaged on a rotated turntable, and hence each object has 72 images. The size of each image is 32×32 pixels with 256 grey levels per pixel. Thus, the original dimension of each image is 1024.

Then, we first randomly select 10 images of each object as the set of training samples, and the rest as the set of testing samples. Further, to make the results more repeatable, we do this operator 20 times randomly. Therefore, the results are obtained in the average sense. By executing six state-of-the-art subspace learning methods and our proposed method, the numerical results are shown in Table 1, and Figures 2 and 3.

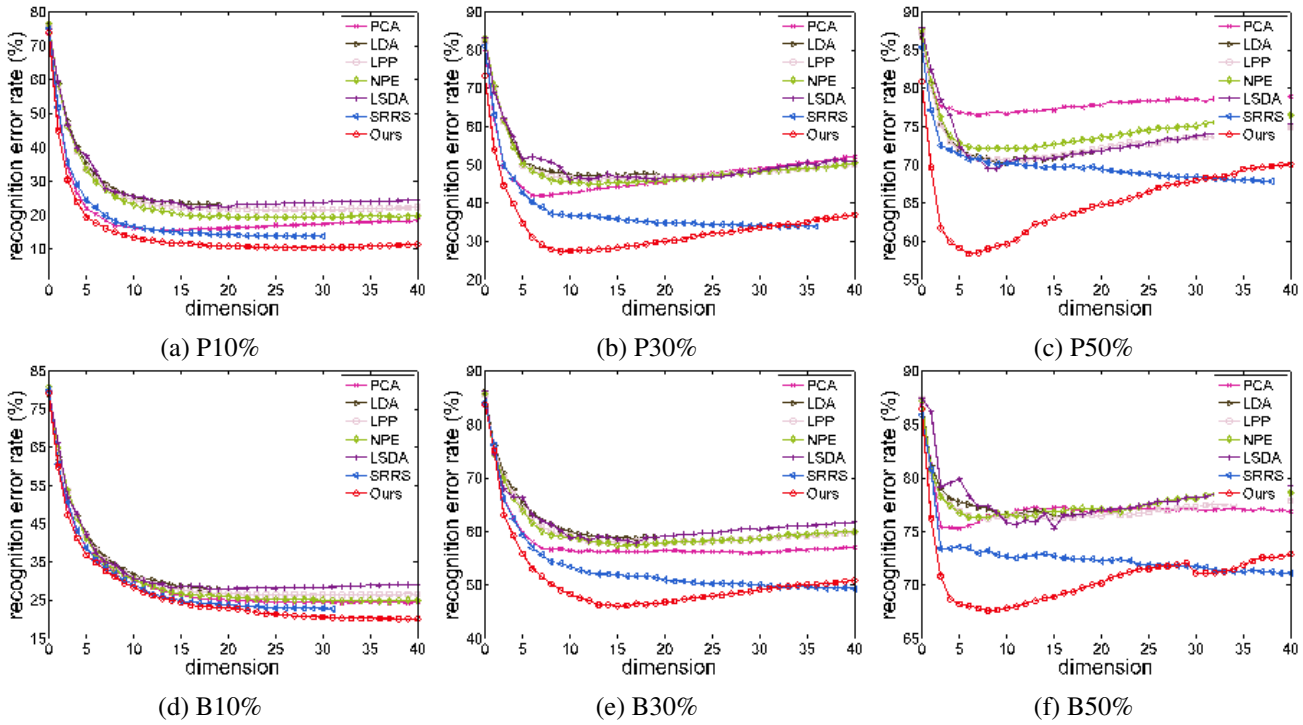
In Table 1, the first column is the data description, where the first letter P and B before the % sign represent pixel corruption and block corruption, respectively. The number in the parentheses means the dimension of the learned subspace. It is seen that the error rate of our proposed method is lower than other six state-of-the-art methods, while the dimension of the learned subspace is low. Specially, the performance is significantly improved in two cases of corruption, and the increment of performance increases with the growth of percentage of corruption. Therefore, our proposed method can obtain better feature representation as well as robustness. This is also reflected in Figure 2. Also, we provide the visualization results with all methods by t-SNE [Maaten and Hinton, 2008] in Figure 3, by which justifies that our proposed method obtains the best result.

4.2 MNIST Digit Dataset

The MNIST dataset of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. The

Method	PCA	LDA	LPP	NPE	LSDA	SRRS	Ours
0%	8.94±0.93(39)	11.75±1.37(16)	10.82±1.28(37)	10.04±1.59(29)	11.04±1.23(16)	8.82±0.92(66)	**8.27±1.86(36)**
P10%	15.38±1.57(14)	22.96±1.21(17)	21.53±1.42(28)	19.30±1.79(25)	22.28±1.16(20)	13.81±1.53(28)	**10.31±1.21(27)**
P20%	25.65±2.07(8)	34.05±1.91(17)	32.96±1.99(16)	30.91±1.48(17)	33.26±1.67(17)	20.61±1.08(33)	**16.84±1.18(15)**
P30%	41.89±1.60(7)	47.10±1.66(10)	46.03±1.47(12)	45.05±2.23(12)	45.97±1.24(10)	33.89±1.42(33)	**27.42±1.66(9)**
P40%	60.92±2.87(5)	58.80±2.40(9)	59.04±1.86(10)	59.44±2.01(9)	57.93±2.01(9)	51.17±1.50(37)	**41.74±1.64(8)**
P50%	76.51±2.18(7)	70.28±1.53(9)	70.39±1.32(11)	72.08±1.38(12)	69.44±1.29(9)	67.84±1.66(38)	**58.35±1.31(6)**
Average	44.07	46.64	45.99	45.36	45.78	37.46	**30.93**
B10%	24.13±1.45(70)	27.87±1.80(19)	26.27±1.70(29)	24.74±1.59(33)	27.21±1.71(19)	22.79±1.61(31)	**20.08±1.28(41)**
B20%	40.20±1.98(23)	41.98±1.22(19)	41.14±1.38(23)	40.59±1.41(26)	41.35±1.04(19)	35.09±1.60(39)	**34.23±1.63(26)**
B30%	56.01±1.35(29)	58.63±1.85(17)	57.81±1.70(20)	57.36±1.72(15)	57.78±1.71(17)	49.02±1.35(46)	**46.07±1.45(14)**
B40%	64.67±1.17(13)	66.33±1.79(18)	66.00±1.47(12)	65.85±1.93(12)	65.24±1.57(18)	58.48±1.48(49)	**54.67±1.78(14)**
B50%	75.27±1.78(5)	76.44±1.49(15)	76.18±1.44(8)	76.30±1.35(7)	75.29±1.09(15)	70.83±1.42(46)	**67.59±1.24(8)**
Average	52.06	54.25	53.48	52.97	53.37	47.24	**44.53**

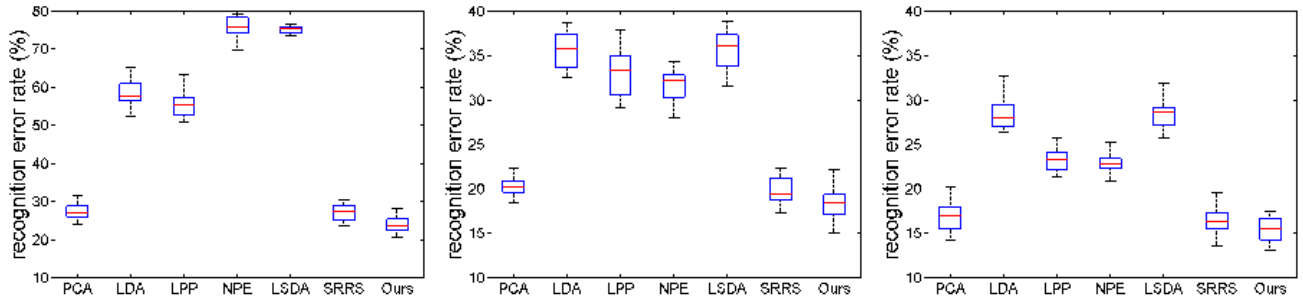
Table 1: The error rates on COIL-100 dataset with different percentages of corruption.



(a) P10% (b) P30% (c) P50%

(d) B10% (e) B30% (f) B50%

Figure 2: Averaged error rates of classification on COIL-100 dataset with different percentages of corruption and dimensions of feature space.



(a) 10 (b) 20 (c) 30

Figure 4: Averaged error rates of classification on MNIST dataset with different training samples.

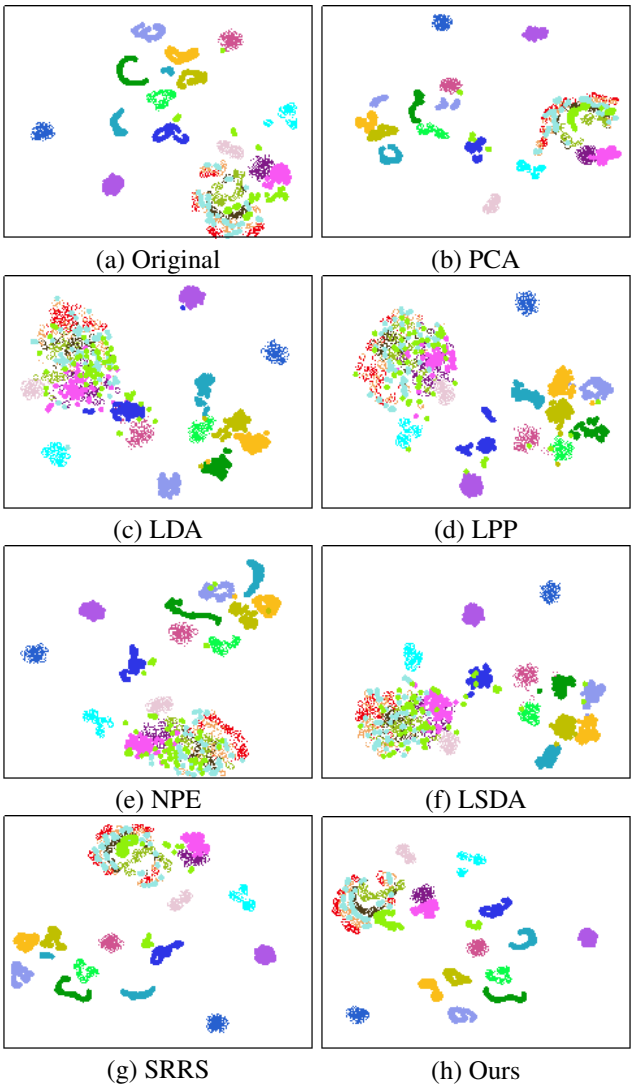


Figure 3: Dataset visualization on two dimensional feature spaces form COIL-100 dataset with 10% pixel corruption. Different colors represent different classes.

digits have been size-normalized and centered in a fixed-size image. The size of each image is 28×28 pixels. Thus, the original dimension of each image is 784. We add 10% pixel corruption, which is the same as the case of COIL-100 dataset.

Then, we first randomly select 1000 images with 10 numbers as our dataset, such that there are 100 images in each group. Then, to test the error rate of classification in the different number of training samples, we randomly select 10, 20, and 30 images of each class as the set of training samples, respectively. Further, to make the results more repeatable, we do this operator 20 times randomly. Therefore, the results are obtained in the average sense. By executing six state-of-the-art subspace learning methods and our proposed method, the statistical results are shown in Figure 4. From Figure 4, we see that the error rate of classification by our proposed method are almost lowest in the different number of training samples.

Method	COIL-100	MNIST
SRRS	1.76	0.63
Ours	0.65	0.58

Table 2: The training time (s) on COIL-100 object dataset and MNIST digit dataset without corruption.

4.3 Discussion

For the computational time, we compare our proposed method with SRRS on COIL-100 object dataset and MNIST digit dataset without corruption. Further, to make the results more repeatable, we conduct this operator 20 times randomly. Therefore, the training time results in the average sense are shown in Table 2, which shows that our proposed method obtains the best classification result with less training time than SRRS.

5 Conclusion

In this paper, we propose an unsupervised subspace learning from a geometric viewpoint. First, we represent the set of all subspaces by the orbit of the rotation group action on the standard subspace. Then, the subspace learning is transferred to a minimization problem on the rotation group. It provides a well geometric explanation of subspace learning. Further, we construct an intrinsic algorithm by applying such geometric structure. Finally, we compare the proposed approach with six state-of-the-art methods on two different kinds of real datasets. The experimental results validate that our proposed method outperforms all compared methods.

Acknowledgments

The research is supported by the National Natural Science Foundation of China under Grant Nos. 11771276, 61573274, 61731009, and the capacity construction project of local universities in Shanghai (18010500600).

References

[Belhumeur *et al.*, 1997] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, 1997.

[Belkin and Niyogi, 2003] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[Cai *et al.*, 2007] Deng Cai, Xiaofei He, Kun Zhou, Jiawei Han, and Hujun Bao. Locality sensitive discriminant analysis. In *IJCAI*, pages 708–713, 2007.

[Candès *et al.*, 2011] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011.

[Chakraborty *et al.*, 2017] Rudrasis Chakraborty, Søren Hauberg, and Baba C. Vemuri. Intrinsic Grassmann

averages for online linear and robust subspace learning. In *CVPR*, pages 6196–6204, 2017.

[Ding and Fu, 2018] Zhengming Ding and Yun Fu. Robust multiview data analysis through collective low-rank subspace. *IEEE Trans. Neural Netw. Lear. Syst.*, 29(5):1986–1997, 2018.

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

[Fang *et al.*, 2018] Xiaozhao Fang, Shaohua Teng, Zhihui Lai, Zhaoshui He, Shengli Xie, and Wai Keung Wong. Robust latent subspace learning for image classification. *IEEE Trans. Neural Netw. Lear. Syst.*, 29(6):2502–2515, 2018.

[Hamm and Lee, 2008] Jihun Hamm and Daniel D. Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *ICML*, pages 376–383, 2008.

[Hauberg *et al.*, 2014] Søren Hauberg, Aasa Feragen, and Michael J. Black. Grassmann averages for scalable robust PCA. In *CVPR*, pages 3810–3817, 2014.

[Hauberg *et al.*, 2016] Søren Hauberg, Aasa Feragen, Raffi Enficiaud, and Michael J. Black. Scalable robust principal component analysis using Grassmann averages. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(11):2298–2311, 2016.

[He *et al.*, 2005a] Xiaofei He, Deng Cai, Shuicheng Yan, and Hong-jiang Zhang. Neighborhood preserving embedding. In *ICCV*, pages 1208–1213, 2005.

[He *et al.*, 2005b] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-jiang Zhang. Face recognition using Laplacianfaces. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(3):328–340, 2005.

[Helgason, 1978] Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*. Academic Press, 1978.

[Huang *et al.*, 2014] Zhiwu Huang, Ruiping Wang, Shiguang Shan, and Xilin Chen. Learning Euclidean-to-Riemannian metric for point-to-set classification. In *CVPR*, pages 1677–1684, 2014.

[Lecun *et al.*, 1998] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Li and Fu, 2016] Sheng Li and Yun Fu. Learning robust and discriminative subspace with low-rank constraints. *IEEE Trans. Neural Netw. Lear. Syst.*, 27(11):2160–2173, 2016.

[Li *et al.*, 2008] Xi Li, Weiming Hu, Zhongfei Zhang, Xiaoqin Zhang, Mingliang Zhu, and Jian Cheng. Visual tracking via incremental Log-Euclidean Riemannian subspace learning. In *CVPR*, pages 1–8, 2008.

[Li *et al.*, 2015] Zechao Li, Jing Liu, Jinhui Tang, and Hanqing Lu. Robust structured subspace learning for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(10):2085–2098, 2015.

[Liu and Yan, 2011] Guangcan Liu and Shuicheng Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *ICCV*, pages 1615–1622, 2011.

[Liu *et al.*, 2010] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *ICML*, pages 663–670, 2010.

[Liu *et al.*, 2013] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):171–184, 2013.

[Maaten and Hinton, 2008] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[Nene *et al.*, 1996] Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (coil 100). *Tech. Rep. CUCS-006-96, Columbia University, New York*, 1996.

[Qiao *et al.*, 2013] Hong Qiao, Peng Zhang, Di Wang, and Bo Zhang. An explicit nonlinear mapping for manifold learning. *IEEE Trans. Cybern.*, 43(1):51–63, 2013.

[Saul and Roweis, 2003] Lawrence K. Saul and Sam T. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003.

[Tenenbaum *et al.*, 2000] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[Turk and Pentland, 1991] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[Vaswani *et al.*, 2018] Namrata Vaswani, Thierry Bouwmans, Sajid Javed, and Praneeth Narayanamurthy. Robust subspace learning: robust PCA, robust subspace tracking, and robust subspace recovery. *IEEE Signal Process. Mag.*, 35(4):32–55, 2018.

[Warner, 2013] Frank W. Warner. Foundations of differentiable manifolds and lie groups. *Springer Science & Business Media*, 2013.

[Wright *et al.*, 2009] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009.

[Zhu *et al.*, 2018] Xiaofeng Zhu, Shichao Zhang, Rongyao Hu, Yonghua Zhu, and Jingkuan Song. Local and global structure preservation for robust unsupervised spectral feature selection. *IEEE Trans. Know. Data Eng.*, 30(3):517–529, 2018.