Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

Subspace Clustering via New Low-Rank Model with Discrete Group Structure Constraint

Feiping Nie, Heng Huang*
Department of Computer Science and Engineering
University of Texas at Arlington
feipingnie@gmail.com, heng@uta.edu

Abstract

We propose a new subspace clustering model to segment data which is drawn from multiple linear or affine subspaces. Unlike the well-known sparse subspace clustering (SSC) and low-rank representation (LRR) which transfer the subspace clustering problem into two steps' algorithm including building the affinity matrix and spectral clustering, our proposed model directly learns the different subspaces' indicator so that low-rank based different groups are obtained clearly. To better approximate the low-rank constraint, we suggest to use Schatten p-norm to relax the rank constraint instead of using trace norm. We tactically avoid the integer programming problem imposed by group indicator constraint to let our algorithm more efficient and scalable. Furthermore, we extend our discussion to the general case in which subspaces don't pass the original point. The new algorithm's convergence is given, and both synthetic and real world datasets demonstrate our proposed model's effectiveness.

1 Introduction

Subspace clustering is an important problem in machine learning, computer vision, such as motion segmentation [Costeira *et al.*, 1997], image compression [Hong *et al.*, 2005] and human face image clustering [Ho *et al.*, 2003]. Subspace clustering assumes that the data points are drawn from multiple low dimensional subspaces. And the goal of subspace clustering is to find such multiple low dimensional subspaces such that all data points are segmented correctly and each group of data points fits into one of the low-dimensional subspaces. Due to the assumption that all data points lie in the multiple subspaces, traditional clustering methods such as K-means, spectral clustering are often difficult to cluster the data points with relatively high accuracy. The subspace clustering is suggested to solve such problem and has gained promising result in the previous research.

There has been a lot of work presenting many different subspace clustering methods. Generally speaking, these methods can be grouped into the following categories: factorization based method such as [Gruber and Weiss, 2004], algebraic based method such as [Vidal *et al.*, 2005], statistic based method such as [Ho *et al.*, 2003], spectral clustering based method such as [Elhamifar and Vidal, 2013]. Recently, sparse learning based methods such as Sparse Subspace Clustering (SSC) [Elhamifar and Vidal, 2013], Subspace Segmentation via Quadratic Programming (SSQP) [Wang *et al.*, 2011] and low-rank based methods such as Robust Subspace Segmentation by Low-Rank Representation (Robust LRR) [Liu *et al.*, 2013], Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction [Liu and Yan, 2011] are all proved to get good performance. For the sparse coding model, it aims to represent one data point with a sparse linear or affine combination of other data points to achieve the subspaces efficiently and naturally. The low-rank based methods utilize low-rank representation opposed to sparse representation to recover the low-dimensional subspaces. The optimal subspace clustering assumptions include different data groups are assigned to different subspaces, there exist enough data points to shape each subspace, and no further subspaces can be derived from the current subspaces. In the other work [Yu and Schuurmans, 2011], the fast closed-form solutions were proposed to solve subspace clustering problem.

We propose a novel low-rank segmentation model to solve the above assumption based subspace clustering problem. Instead of learning the affinity matrix done by previous subspace clustering methods including SSC [Elhamifar and Vidal, 2013] and LRR [Liu *et al.*, 2013], we propose to learn the group indicator directly such that the low-rank structure for each subspace is clear. We utilize the Schatten p-norm instead of trace norm to relax the rank constraint to better approximate the low-rank constraint, tactically avoid the integer programming problem imposed by the group indicator constraint, and extend our new model to discuss the general case in which the subspaces don't pass the original point. An efficient method is suggested with convergence prove to solve the proposed subspace segmentation model. Both synthetic and real world data sets demonstrate our proposed new subspace clustering model's effectiveness.

2 New Subspace Clustering Model

To begin with, we denote the groups for which data points in $X \in \mathbb{R}^{d \times n}$ are clustered as $\{X_1, X_2, ..., X_k\}$. Here we first

*To whom all correspondence should be addressed. This work was partially supported by US NSF-IIS 1117965, NSF-IIS 1302675, NSF-IIS 1344152, NSF-DBI 1356628, NIH R01 AG049371.

discuss the condition when each subspace passes through the origin. We will talk about the case when each subspace might not pass through the origin later.

In order to get the optimal subspace clustering result, we need seek the low-rank subspaces to accommodate different groups of data points. To achieve the optimal low-rank solutions, we can minimize the rank of all subspace simultaneously as:

$$\min \sum_{i=1}^{k} rank(X_i)\,. \tag{1}$$

However, the above objective often has trivial solution. If all data points are allocated into a single subspace, the objective value in Eq. (1) is $\min(d, n)$. If the ground-truth are k groups, denote $r_i = rank(X_i)$, the objective value $\sum_{i=1}^{k} r_i$ is often larger than $\min(d, n)$. For example, there are 5 subspaces of dimension 3 in $\mathbb{R}^{10}$, and 20 points per subspace are sampled. As a result, each data matrix X_i is 10 by 20 and of rank 3. The objective value is 15 (sum of ground-truth ranks). But, if all data points are allocated into a single subspace, the objective value is 10, which is even smaller than that of the ground-truth. Thus, the objective function in Eq. (1) is not suitable for the subspace clustering task.

Because in subspace clustering, $r_i << \min(d, n)$, the value of $\sum_{i=1}^{k} r_i^2$ is smaller than $(\min(d, n))^2$ in general cases. For example, in the above example, the sum of the square ranks $\sum_{i=1}^{k} r_i^2 = 45$. If all data points are in a single subspace, $(\min(d, n))^2 = 10^2 = 100 > 45$. Thus, we can avoid the trivial solution. Based on this consideration, we propose to solve the following objective function:

$$\min \sum_{i=1}^{k} (rank(X_i))^2\,. \tag{2}$$

We denote a cluster indicator matrix by $Idx \in \mathbb{R}^{k\times n}$, $Idx(i, j) = 1$ if the j-th data point is grouped into the i-th cluster or subspace, and $Idx(i, j) = 0$ otherwise. The k diagonal matrices $G_1, G_2, ..., G_k$ are defined, where the diagonal elements of $G_i (1 \le i \le k)$ are formed by the i-th row of matrix Idx. Identity matrix is denoted by I. Note that $rank(X_i) = rank(XG_i)$, the problem (2) is equivalent to the following problem:

$$\begin{aligned} &\min_{G_i|_{i=1}^{k}} \sum_{i=1}^{k} (rank(XG_i))^2 \\ &s.t. \quad G_i|_{i=1}^{k} \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{3}$$

where the summation constraint ensures each data point is assigned to one and only one group.

The objective function is non-convex, so does the discrete constraint. It is well known that trace norm (denote as $\|\cdot\|_*$) is the convex envelope of the rank function, and is thought to be the best convex approximation [Recht *et al.*, 2010]. The discrete constraint for the group indication matrix $G_i|_{i=1}^{k} \subseteq \{0, 1\}^{n\times n}$ can be relaxed to be $G_i|_{i=1}^{k} \subseteq [0, 1]^{n\times n}$ which is explained as probabilities for each data point belonging to certain group. As a result, based on the convex relaxation of both objective function and constraint, the problem (3) now becomes:

$$\begin{aligned} &\min_{G_i|_{i=1}^{k}} \sum_{i=1}^{k} \|XG_i\|_*^2 \\ &s.t. \quad G_i|_{i=1}^{k} \subseteq [0,1]^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{4}$$

But such convex relaxation format may not be the best approximation to the original non-convex problem (3), because it includes two relaxation including trace norm and continuous constraint. So we want to use other relaxation which can approximate the original problem (3) better. The Schatten p-norm is a better relaxation for the rank objective function than the trace norm when $p < 1$ based on the definition [Nie *et al.*, 2012; 2015]. As a result, in this paper we relax the rank based objective function to the following one with binary constraint for each $G_i|_{i=1}^{k}$:

$$\begin{aligned} &\min_{G_i|_{i=1}^{k}} \sum_{i} (\|XG_i\|_{S_p}^p)^2 \\ &s.t. \quad G_i|_{i=1}^{k} \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{5}$$

where $\|F\|_{S_p}^p = Tr((FF^T)^{\frac{p}{2}})$ is power p of the Schatten p-norm of matrix F.

In the next section, an iteration based re-weighted method [Nie *et al.*, 2010; 2012; 2014] will be applied to solve it with convergence analysis.

3 Optimization Algorithm

According to the definition of Lagrange function, the Lagrangian function of problem (5) is

$$L(G_i|_{i=1}^{k}, \Lambda) = \sum_i (\|XG_i\|_{S_p}^p)^2 + g(\Lambda, G_i|_{i=1}^{k})$$

where $g(\Lambda, G_i|_{i=1}^{k})$ encodes the constraints in problem (5).

Taking derivative w.r.t. $G_i|_{i=1}^{k}$ and setting to zero, it becomes

$$\sum_i 2X^T D_i X G_i + \frac{\partial g(\Lambda, G_i|_{i=1}^{k})}{\partial\, G_i|_{i=1}^{k}} = 0 \tag{6}$$

where

$$D_i = p\,\|XG_i\|_{S_p}^p \,(XG_i^2X^T)^{\frac{p-2}{2}} \tag{7}$$

Note that D_i depends on G_i, we use an iteration based algorithm to get the solution that satisfies Eq. (6). First, according to Eq. (7), $D_i|_{i=1}^{k}$ can be calculated based on the current solution $G_i|_{i=1}^{k}$. If $D_i|_{i=1}^{k}$ is given, the optimal solution $G_i|_{i=1}^{k}$ to the following problem will satisfy Eq. (6):

$$\begin{aligned} &\min_{G_i|_{i=1}^{k}} \sum_{i} Tr(G_i^T X^T D_i X G_i) \\ &s.t. \quad G_i|_{i=1}^{k} \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{8}$$

Then update the current solution $G_i|_{i=1}^{k}$ by the optimal solution to the problem (8). It can be proved that this iterative procedure converges to a local optimum of the problem (5). Here is the intuition on the changes from Eq. (6) to Eq. (8). Eq. (6) which has the Lagrange multipliers is one of the KKT conditions w.r.t. the derivative of primal variables G_i, and after that we set D_i fixed. The same KKT condition w.r.t. the derivative of primal variables G_i is used to Eq. (8). The objective derivative w.r.t. G_i in Eq. (8) is the same as the first part of Eq. (6). Because our constraint does not change, it has the same form as the second part for multipliers in Eq. (6). The reason why we introduce the multipliers in Eq. (6) is to show the connection between Eq. (6) and Eq. (8). Because the constraints are not changed and we want to use iteration based re-weighted method to solve this optimization problem instead of KKT condition involved method, the multipliers disappear.

Problem (8) can be rewritten as

$$\begin{aligned} &\min_{G_i|_{i=1}^k} \sum_i Tr(A_i G_i^2) \\ &s.t. \quad G_i|_{i=1}^k \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^k G_i = I \end{aligned} \tag{9}$$

where $A_i = X^T D_i X$. the problem (9) can be further changed to the following problem because of the discrete constraint $G_i|_{i=1}^k \subseteq \{0,1\}^{n\times n}$:

$$\begin{aligned} &\min_{G_i|_{i=1}^k} \sum_i Tr(A_i G_i) \\ &s.t. \quad G_i|_{i=1}^k \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^k G_i = I \end{aligned} \tag{10}$$

Note that $G_i|_{i=1}^k$ are n by n diagonal matrices, the problem (9) can be rewritten as the following problem:

$$\min_{g_{ci}\in\{0,1\}, \sum_{i=1}^k g_{ci}=1} \sum_{i=1}^k \sum_{i=c}^n a_{ci} g_{ci} \tag{11}$$

where g_{ci} is the c-th diagonal element of matrix G_i and a_{ci} is the c-th diagonal element of matrix A_i.

The optimal solution to the problem (11) can be easily obtained as follows

$$g_{ci} = \begin{cases} 1, & i = \arg\min_l a_{cl} \\ 0, & otherwise \end{cases} \tag{12}$$

The algorithm to solve problem (5) is summarized in Algorithm 1. As can be seen, the algorithm is very simple and concise.

Algorithm 1 Algorithm to solve problem (5).

Initialize $G_i|_{i=1}^k$ such that the constraints in the problem (5) are satisfied
while not converge **do**
1. Calculate $D_i|_{i=1}^k = p \left\|XG_i\right\|_{S_p}^p (XG_i^2X^T)^{\frac{p-2}{2}}$
2. Calculate $A_i|_{i=1}^k = X^T D_i X$
3. Update $G_i|_{i=1}^k$, where the c-th diagonal element g_{ci} of G_i is updated by Eq. (12)

end while

3.1 Convergence Analysis

We have the following theorem on the convergence of the proposed algorithm.

Theorem 1 *The Algorithm 1 will finally converge to a local optimal solution of the problem (5) when* $p = 1$.

Proof: Assume the updated G_i in step 3 of Alg. 1 is $\tilde{G}_i$, Since $\tilde{G}_i$ is the optimal solution to the problem (8), we have

$$\sum_i Tr(\tilde{G}_i^T X^T D_i X \tilde{G}_i) \le \sum_i Tr(G_i^T X^T D_i X G_i)$$

which can be written as

$$\sum_i Tr(D_i X \tilde{G}_i^2 X^T) \le \sum_i Tr(D_i X G_i^2 X^T) \tag{13}$$

According to step 1 in Alg. 1, Eq. (13) can be written as

$$\begin{aligned} &\sum_i \|XG_i\|_{S_p}^p \, Tr((XG_i^2X^T)^{\frac{p-2}{2}} X\tilde{G}_i^2X^T) \\ &\le \sum_i (\|XG_i\|_{S_p}^p)^2 \end{aligned} \tag{14}$$

According to the Cauchy-Schwarz inequality, it can be proved that, when $p = 1$ we have

$$\begin{aligned} &\sum_i (\left\|X\tilde{G}_i\right\|_{S_p}^p)^2 \le \\ &\sum_i Tr((XG_i^2X^T)^{\frac{p}{2}}) Tr((XG_i^2X^T)^{\frac{p-2}{2}} X\tilde{G}_i^2X^T) \end{aligned} \tag{15}$$

Therefore, based on Eqs. (14-15), we have

$$\sum_i (\left\|X\tilde{G}_i\right\|_{S_p}^p)^2 \le \sum_i (\|XG_i\|_{S_p}^p)^2 \tag{16}$$

Eq. (16) indicates that the objective function in the problem (5) will monotonically decrease during the iteration until the algorithm converges. Since the objective in problem (5) has clearly lower bound 0, the algorithm will converge. When it converges, we can see that Eq. (6) will always be satisfied, so the algorithm will converge to a local optimum solution to the problem (5) when $p = 1$.

$\square$

In practice, we observed the algorithm is also converged for $0 < p < 1$. If the objective in problem (5) is changed to $\sum_i (\|XG_i\|_{S_p}^p)^m (m > 2)$ (the D_i in step 1 is changed to $\frac{mp}{2}(\|XG_i\|_{S_p}^p)^{m-1}(XG_i^2X^T)^{\frac{p-2}{2}}$ accordingly), and/or the constraint is changed to the one as in Eq.(4) (the Eqs. (11-12) are changed accordingly), the convergence is also observed[1].

3.2 Complexity Analysis

We suppose $d < n$ in the following analysis. In step 1, we need to compute $D_i|_{i=1}^k$. Computing $\|XG_i\|_{S_p}^p$ need SVD of XG_i, which takes $O(nd^2)$. Suppose the SVD of XG_i is $XG_i = U\Sigma V^T$, then $(XG_i^2X^T)^{\frac{p-2}{2}} = U\Sigma^{p-2}U^T$, which

[1] $G_i|_{i=1}^k = \frac{1}{n}$ is a trivial optimal solution to problem (4) because of the following facts: problem (4) is convex, Alg. 1 will converge to a local minimum, and the solution will be unchanged with this trivial solution as initialization in Alg. 1.

takes $O(d^3)$. So computing D_i takes $O(d^2n)$ and computing $D_i|_{i=1}^k$ takes $O(d^2nk)$. Step 2 only needs to compute the diagonal elements of A_i, so step 2 takes $O(d^2nk)$. Computing $G_i|_{i=1}^k$ in step 3 takes $O(nk)$. In summary, the computational complexity of Algorithm 1 is $O(d^2nkt)$, where t is the iteration number. In the experiments we found the algorithm converges very fast, and always converges in 5-10 iterations. According to the above analysis, the time consumption of Algorithm 1 is linear *w.r.t.* the number of data. Therefore, the algorithm can easily handle large scale data if the dimensionality of data is not too high.

4 Subspace Clustering without Passing through Original Point

In the previous section, we assume that each subspace passes through the origin. If a subspace does not pass through the origin, the rank of the group will reduce one as long as the subspace is shifted to pass through the origin, as can be seen in Fig. 1. Therefore, the shifts can also be learned such that the ranks of the groups are minimized. In this case, instead of minimizing Eq. (2), we minimize the following objective function:

$$\sum_{i=1}^{k}(rank(X_i - u_i\mathbf{1}^T))^2 \tag{17}$$

where $\mathbf{1}$ denotes a column vector with all ones. Similar to the previous section, minimizing Eq. (17) is equivalent to solving the following problem:

$$\begin{aligned} &\min_{u_i, G_i|_{i=1}^k} \sum_{i=1}^{k}(rank((X - u_i\mathbf{1}^T)G_i))^2 \\ &s.t. \quad G_i|_{i=1}^k \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{18}$$

Similarly, the following problem for the subspace clustering problem in this case is proposed to solve:

$$\begin{aligned} &\min_{u_i, G_i|_{i=1}^k} \sum_{i}(\left\|(X - u_i\mathbf{1}^T)G_i\right\|_{S_p}^p)^2 \\ &s.t. \quad G_i|_{i=1}^k \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{19}$$

This problem can also be efficiently solved by iteration based re-weighted algorithm. We define $X_{u_i} = X - u_i\mathbf{1}^T$. Similarly to Eq. (8), we need to solve the following problem:

$$\begin{aligned} &\min_{u_i, G_i|_{i=1}^k} \sum_{i} Tr(G_i^T X_{u_i}^T D_i X_{u_i} G_i) \\ &s.t. \quad G_i|_{i=1}^k \subseteq \{0,1\}^{n\times n}, \sum_{i=1}^{k} G_i = I \end{aligned} \tag{20}$$

where

$$D_i = p\left\|X_{u_i}G_i\right\|_{S_p}^p (X_{u_i}G_i^2X_{u_i}^T)^{\frac{p-2}{2}} \tag{21}$$

When fix $u_i|_{i=1}^k$, the $G_i|_{i=1}^k$ can be solved as in Eqs. (11-12) with $A_i = X_{u_i}^T D_i X_{u_i}$.

When fix $G_i|_{i=1}^k$, the problem (20) becomes

$$\min_{u_i} Tr(G_i^T(X - u_i\mathbf{1}^T)^T D_i (X - u_i\mathbf{1}^T)G_i) \tag{22}$$

The optimal solution to the problem (22) can be easily derived as

$$u_i = \frac{1}{\mathbf{1}^T G_i^2 \mathbf{1}} X G_i^2 \mathbf{1}$$

The algorithm to solve problem (19) is summarized in Algorithm 2.

Algorithm 2 Algorithm to solve problem (19).

Initialize $G_i|_{i=1}^k$ such that the constraints in the problem (19) are satisfied. Initialize $u_i|_{i=1}^k = \mathbf{0}$.
while not converge **do**
1. Calculate $X_{u_i}|_{i=1}^k = X - u_i\mathbf{1}^T$
2. Calculate $D_i|_1^k = p\left\|X_{u_i}G_i\right\|_{S_p}^p (X_{u_i}G_i^2X_{u_i}^T)^{\frac{p-2}{2}}$
3. Calculate $A_i|_{i=1}^k = X_{u_i}^T D_i X_{u_i}$
4. Update $G_i|_{i=1}^k$, where the c-th diagonal element g_{ci} of G_i is updated by Eq. (12)
5. Update $u_i|_{i=1}^k = \frac{1}{\mathbf{1}^T G_i^2 \mathbf{1}} X G_i^2 \mathbf{1}$

end while

5 Experiments

5.1 Experimental Results on Synthetic Data

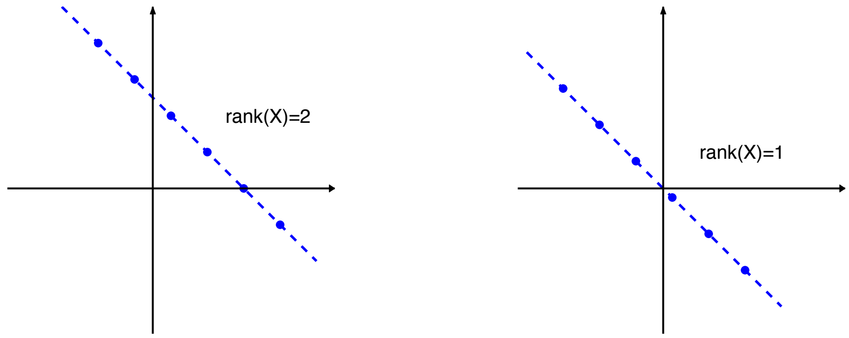
First, we test our algorithm on three challenge Synthetic datasets, each of which comprises several groups distributed on different lower dimensional subspaces but with 5% level noises to deviate from the subspaces. These subspace intersect each other, so it is a very difficult subspace clustering problem.

We run our algorithm with ten different initializations and select the results with the best objective values. The results are shown in Fig. 2, which we can see that our algorithm can correctly find the low-dimensional structure and find the correct groups on these challenge datasets. The results indicate that the proposed method can effectively solve the subspace clustering problem and has the power to find the low-dimensional structure information hidden in data, which is also the goal in Generalized Principal Component Analysis (GPCA) [Vidal *et al.*, 2005].

We also test our algorithm on a synthetic high-dimensional dataset. In this dataset, we randomly generate 250 data points with 50 dimensions lying in 5 different subspaces. Each subspace has 50 data points, and the dimensions of the 5subspaces are $5, 10, 15, 20, 25$, respectively. In the data, we also add 5% level noises to deviate from the subspaces. The nearest neighbor graph of the data based on Euclidean distance is shown in Fig. 3(a), and the result $Idx'Idx$ of the proposed method is shown in Fig. 3(a), where Idx is defined in Section 2. This result also verify that the proposed method is an effective subspace clustering method.

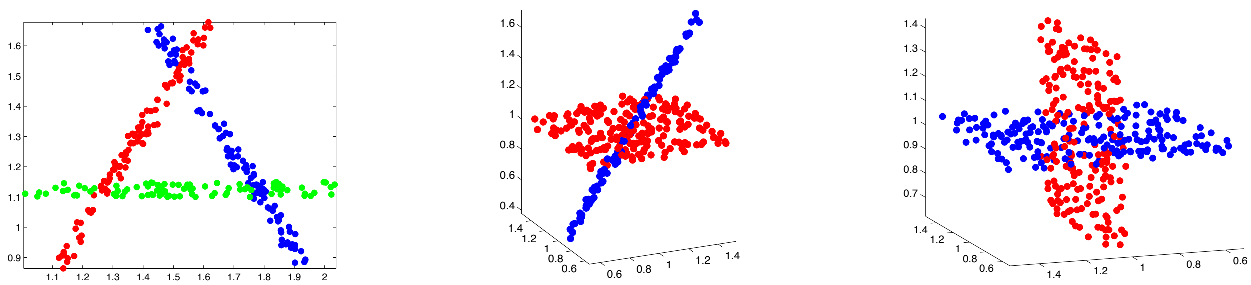
5.2 Experiments on Real World Data

Here we just evaluate Algorithm 2, and leave out the Algorithm 1, because Algorithm 2 is the subspace clustering in the case which has more wide application. We evaluate the algorithm 2's performance in handling two real-world problems: motion segmentation and human images' clustering.



(a) (b)

Figure 1: When the six data points are shifted such that the subspace pass through the origin, then the rank of the data matrix with these data points is reduced from 2 to 1.



(a) (b) (c)

Figure 2: Results of the proposed subspace clustering method on three synthetic subspace segmentation problems.

Motion Segmentation in Hopkins 155 Dataset

We first test our model on the Hopkins 155 motion dataset. This dataset consists of 155 sequences, which have about 39-550 points tracked from 2 or 3 motions, separately. And in these 155 datasets, there are 120 datasets having two motions and 35 datasets having three motions. Our goal is to cluster these points into groups according to their motions in each sequence. Each sequence is a dataset, so there are 155 datasets with different properties on subspace's number, data samples' number. There are about 10 sequences that are grossly corrupted, having high error levels, although most of the outliers in the datasets have been removed manually.

We combine PCA and K-means methods to initialize our $G_i (1 \leq i \leq k)$. We use the PCA to project the coordinates in each sequence into the dimensions ranging from 5 to 20. Then we use K-means method to get our initialized $G_i (1 \leq i \leq k)$. We compare the performance of our algorithm with the best state-of-the-art motion segmentation algorithms: Sparse Subspace Clustering (SSC) [Elhamifar and Vidal, 2013], Subspace Segmentation via Quadratic Programming (SSQP) [Wang *et al.*, 2011], Robust Low-Rank Representation (Robust LRR) [Liu *et al.*, 2013], Groupwise Constrained Reconstruction (GCR) [Li *et al.*, 2012], Groupwise Constrained Reconstruction with Dirichlet Process (GCR-DP) [Li *et al.*, 2012]. The final result with the known best results for each algorithm are reported in Table 2.

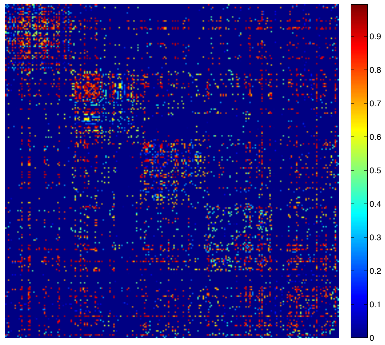
Category	Checkerboard	Traffic	Others	All
No. of Seq.	78	31	11	120
Points	291	241	155	266
Frames	28	30	40	30
No. of Seq.	26	7	2	35
Points	437	332	122	398
Frames	28	31	31	29

Table 1: Hopkins 155 dataset description

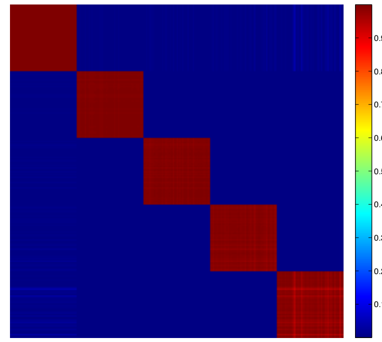
From the Table 2 we can observe that: Our proposed method outperforms other state-of-art methods in all accuracy and error's measurement including mean accuracy, minimum accuracy and standard deviation of error. The main reason is that our method takes the Schatten p- norm into account, which is the relaxation that better approximates the original problem (3) than the trace norm approximation when $p < 1$. Also, our method has considered the condition when the subspace does not pass through the origin to learn the shifts such that the ranks of the groups are minimized.

Subspace Clustering on Real-World Image Datasets

In order to check our method's suitability in the real human images' clustering condition, face datasets including JAFFE



(a) Nearest neighbor graph



(b) Learned structure

Figure 3: Left: The nearest neighbor graph of a synthetic high-dimensional dataset. Right: the result $Idx'Idx$ of the proposed method.

METHOD	Mean	Median	Min	Std.
SSC [Elhamifar and Vidal, 2013]	97.29	100	57.66	-
SSQP [Wang *et al.*, 2011]	95.36	100	54.50	-
GCR-DP [Li *et al.*, 2012]	97.67	100	55.32	-
GCR [Li *et al.*, 2012]	96.08	99.70	58.33	-
Robust LRR [Liu *et al.*, 2013]	98.29	-	-	4.85
Our method	**98.59**	100	**58.63**	**4.76**

Table 2: Subspace clustering accuracy(%) and standard error(%) on the Hopkins 155 Dataset

METHOD	JAFFE	XM2VTS	MSRA	PALM
Robust LRR	95.31	58.56	71.84	80.75
Convex LRR	95.31	56.80	70.65	85.10
SSC	95.31	57.32	73.95	81.24
PCA	93.24	56.19	62.42	75.65
SSQP	96.71	61.27	54.20	78.20
Our method	**96.71**	**61.27**	**93.11**	**85.75**

Table 3: Subspace clustering accuracy(%) comparisons

[Lyons *et al.*, 1998], MSRA [Liu *et al.*, 2007], XM2VTS [XM2,] and another human palm image dataset called PALM [Yan *et al.*, 2007] are used.

We include Robust LRR [Liu *et al.*, 2013], Convex LRR [Liu *et al.*, 2013], its very similar model SSQP [Wang *et al.*, 2011], SSC [Elhamifar and Vidal, 2013], and the classic Principle Component Analysis (PCA) to be our comparatives in the human images' clustering.

Similar to the motion segmentation method, we also combine PCA and K-means methods to initialize our $G_i(1 \leq i \leq k)$. We use the PCA to project the coordinates in each dataset into the dimensions ranging from 10 to 50. Because the dataset for MSRA, XM2VTS, and PALM are relatively large, we first use PCA to project the original data samples to a low dimension based on the PCA ratio equaling to 0.98 to facilitate the whole progress. The best results for each methods can been seen on Table 3.

From Table 3 we can find that: our proposed method performs better than other methods in four datasets, which proves that our method is more suitable for the subspace clustering problem than other proposed subspace clustering methods. Note that in the MSRA dataset, our method's clustering accuracy is 93.11 %, which is the best result so far to the best of our knowledge. It is because the MSRA face samples are more satisfied to our proposed subspace clustering model than other 3 human image samples.

6 Conclusions

In this paper, a new subspace clustering model based on the Schatten p-norm is proposed, and it is a better approximation to the rank minimization problem than the trace norm approximation and others. According to the subspace model we propose, an iteration based re-weighted method with both effectiveness and efficiency is suggested to solve this model. The convergence of our proposed algorithm is proved. What's more, we consider the condition when the subspace does not pass through the origin to learn the shifts such that the ranks of the groups are minimized. Both the Synthetic datasets and Real world datasets demonstrate our method's effectiveness in dealing with subspace clustering problems.

References

[Costeira *et al.*, 1997] Joo Costeira, Takeo Kanade, and Mutin Analysis Invariants. A multi-body factorization method for independently moving objects. *International Journal of Computer Vision*, 29:159–179, 1997.

[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.

[Gruber and Weiss, 2004] Amit Gruber and Yair Weiss. Multibody factorization with uncertainty and missing data using the em algorithm. In *CVPR (1)*, pages 707–714, 2004.

[Ho *et al.*, 2003] Jeffrey Ho, Ming-Hsuan Yang, Jongwoo Lim, Kuang-Chih Lee, and David Kriegman. Clustering appearances of objects under varying illumination conditions. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR'03, pages 11–18, Washington, DC, USA, 2003. IEEE Computer Society.

[Hong *et al.*, 2005] Wei Hong, John Wright, Kun Huang, and Yi Ma. A multi-scale hybrid linear model for lossy image representation. *Computer Vision, IEEE International Conference on*, 1:764–771, 2005.

[Li *et al.*, 2012] Ruijiang Li, Bin Li, Cheng Jin, and Xiangyang Xue. Groupwise constrained reconstruction for subspace clustering. In *ICML*, 2012.

[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.*, 2007] Tie Liu, Jian Sun, Nan ning Zheng, Xiaoou Tang, and Heung yeung Shum. Learning to detect a salient object. In *In CVPR*, 2007.

[Liu *et al.*, 2013] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

[Lyons *et al.*, 1998] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, FG '98, pages 200–, Washington, DC, USA, 1998. IEEE Computer Society.

[Nie *et al.*, 2010] Feiping Nie, Heng Huang, Xiao Cai, and Chris Ding. Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization. In *NIPS*, 2010.

[Nie *et al.*, 2012] Feiping Nie, Heng Huang, and Chris H. Q. Ding. Low-rank matrix recovery via efficient schatten p-norm minimization. In *AAAI*, 2012.

[Nie *et al.*, 2014] Feiping Nie, Jianjun Yuan, and Heng Huang. Optimal mean robust principal component analysis. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pages 1062–1070, 2014.

[Nie *et al.*, 2015] 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.

[Recht *et al.*, 2010] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[Vidal *et al.*, 2005] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1945–1959, 2005.

[Wang *et al.*, 2011] Shusen Wang, Xiaotong Yuan, Tiansheng Yao, Shuicheng Yan, and Jialie Shen. Efficient subspace segmentation via quadratic programming. In *AAAI*, 2011.

[XM2,] http://www.ee.surrey.ac.uk/cvssp/xm2vtsdb/.

[Yan *et al.*, 2007] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(1):40–51, January 2007.

[Yu and Schuurmans, 2011] Yaoliang Yu and Dale Schuurmans. Rank/norm regularization with closed-form solutions: Application to subspace clustering. In *UAI*, pages 778–785, 2011.