Robust Subspace Clustering by Logarithmic Hyperbolic Cosine Function

Lei Cao, Long Shi, *Member, IEEE*, Jun Wang, Zhendong Yang, and Badong Chen, *Senior Member, IEEE*

***Abstract*—As an important category of clustering methods, subspace clustering algorithms have arisen particular attention during the last decade. Most subspace clustering algorithms are designed by first constructing a similarity matrix and then using spectral clustering algorithms to perform clustering. How to learn a suitable representation matrix to construct the similarity matrix is essential to the clustering performance. In most existing algorithms, the representation matrix is solved by norm-minimization, which commonly enforces the error matrix with nuclear norm or sparsity norm. However, these methods may fail to achieve satisfactory performance for real data contaminated by complex noise. To this end, we propose a novel robust subspace clustering method based on the Logarithmic Hyperbolic Cosine Function (LHCF). We theoretically analyze the grouping effect, as well as the convergence behavior, which illustrates that highly correlated samples can be grouped into the same cluster. Experimental results conducted on the Extended Yale B dataset show that the newly proposed algorithm yields better clustering performance compared with some advanced methods.**

***Index Terms*—Complex noise, convergence behavior, grouping effect, logarithmic cosine function, subspace clustering.**

I. Introduction

In MACHINE learning, subspace clustering technique has been rapidly developed and widely used in various scenarios over the last decade, e.g., face clustering [1], [2], motion segmentation [3], [4], and character recognition [5], [6]. Subspace clustering aims to recover the underlying low-dimensional subspace structures and then divide samples into the corresponding groups [7], [8]. Subspace clustering methods mainly involve iterative methods [9], [10], algebraic methods [11], [12], statistical methods [13], [14], and spectral clustering-based methods [15], [16], among which spectral clustering-based methods have received particular attention due to their satisfactory clustering performance even in the case of lacking some priori information. Basically, we can implement spectral clustering algorithms by two steps [17], [18], [19]: 1) learn the representation matrix to construct a similarity matrix; 2) perform a spectral clustering algorithm on the similarity matrix to find clusters.

In [20], the author first studied the Sparse Subspace Clustering (SSC) algorithm based on a fundamental observation concerning sparse representation of data points, but SSC is not able to exploit the correlation information of data from the same subspace. Aimed at recovering the subspace structures from unclean data, the Low-rank Representation (LRR) algorithm was put forward in [21] by seeking the *lowest rank* representation, which takes into account the correlation information of data. Essentially, SSC and LRR can be distinguished from the regularization on representation matrix: SSC enforces the representation matrix with sparse norm while LRR employs the nuclear norm.

It is important to note that effectively utilizing the block diagonal property can significantly improve the accuracy of clustering results [22]. In order to ensure the block diagonal property between-cluster as well as the within-cluster affinities, the Least Squares Regression (LSR) algorithm was proposed with Frobenius norm enforced on its representation matrix [23], [24]. To balance SSC and LSR, the author used trace Lasso to propose the Correlation Adaptive Subspace Segmentation (CASS) algorithm [25]. CASS is able to simultaneously execute data selection and cluster correlated data, thanks to the *adaptive* characteristic of trace Lasso. In practical scenarios, the sampling data are often contaminated by complex noise, thus the above mentioned methods may suffer from performance degradation. In [26], the author proposed a Robust Subspace Segmentation (RSS) scheme based on establishing a framework to find the representation of data and the similarity matrix in a unified way, which improves the robustness against heavily noisy data. In [27], the Cauchy Loss Function (CLF) was used for designing a robust subspace clustering scheme.

In this letter, we focus on the investigation concerning robust subspace clustering. Our contributions mainly include:

- A novel subspace clustering algorithm based on the Logarithmic Hyperbolic Cosine Function (LHCF) is proposed. Compared with the commonly used l_1 loss, the influence function of LHCF (it first derivative) is bounded without discontinuities [28], which makes it more suitable for gradient-based optimization algorithms as well as improves the model robustness. Thus, LHCF is expected to be a good candidate to reduce the negative effects of complex noise.

Manuscript received 17 January 2023; revised 18 April 2023; accepted 20 April 2023. Date of publication 1 May 2023; date of current version 5 May 2023. The work of Long Shi was supported by the National Natural Science Foundation of China under Grant 62201475. The work of Badong Chen was supported by the National Natural Science Foundation of China under Grants U21A20485 and 61976175. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Saurabh Prasad. *(Corresponding author: Long Shi.)*

Lei Cao, Long Shi, and Zhendong Yang are with the School of Computing and Artificial Intelligence, Financial Intelligence and Financial Engineering Key Laboratory of Sichuan Province, Southwestern University of Finance and Economics, Chengdu 611130, China (e-mail: caolei2000@smail.swufe.edu.cn; shilong@swufe.edu.cn; yzd123@smail.swufe.edu.cn).

Jun Wang is with the School of Management Science and Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China (e-mail: wangjun1987@swufe.edu.cn).

Badong Chen is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: chenbd@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/LSP.2023.3271828

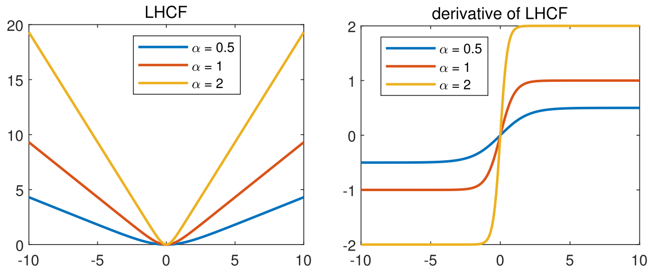

Fig. 1. Graphical curves of LHCF and its derivative for different α.

- The grouping effect as well as the convergence behavior of the newly proposed method is theoretically analyzed.
- Experiments carried out on the Extended Yale B dataset illustrate that the new method provides more favorable clustering performance than some advanced algorithms.

II. Robust Subspace Clustering by LHCF

In this section, we briefly review the fundamental concept of subspace clustering. Then, we introduce LHCF. Finally, the detailed derivation of the algorithm is presented.

A. Preliminary

Let $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_k] = [\mathbf{x_1}, \ldots, \mathbf{x_n}] \in \mathbb{R}^{d\times n}$ *denote a set of data vectors draw from a union of* k *linear subspaces* $\{S_i\}_{i=1}^{k}$*, where* $\mathbf{X}_i$ *is a collection of* n_i *data vectors drawn from the subspace* S_i*, and* $n = \sum_{i=1}^{k} n_i$*. The goal of subspace clustering is to segment the data according to the underlying subspaces they are drawn from.*

Generally, spectral clustering-based subspace clustering algorithms can be derived by solving the following optimization problem

$$\min_{\mathbf{Z},\mathbf{E}} f(\mathbf{Z}) + g(\mathbf{E}) \quad \text{s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \tag{1}$$

where $\mathbf{X}$ is the raw data matrix, $\mathbf{Z}$ is the representation matrix to be learned, $\mathbf{E}$ is the noise-induced matrix. The functions $f(\mathbf{Z})$ and $g(\mathbf{E})$ are used to restrict $\mathbf{Z}$ and $\mathbf{E}$, respectively. In SSC, l_1-norm is used to regularize $\mathbf{Z}$, while in LRR, nuclear-norm is enforced on $\mathbf{Z}$.

B. Logarithmic Hyperbolic Cosine Function

We consider a robust function called LHCF, which is defined by

$$g(x) = \ln[\cosh(\alpha x)], \tag{2}$$

where α is a scaling factor. The first derivative of LHCF is given by

$$g'(x) = \alpha \tanh(\alpha x). \tag{3}$$

Fig. 1 shows the graphical curves of LHCF and its derivative for different α. It is seen that the derivative of LHCF is bounded such that the large outliers can be effectively suppressed. As a result, even in the presence of large noise, LHCF is capable of providing good robustness. Moreover, changing the value of α results in a range of distinct suppression effects.

Algorithm 1: Learn the Representation Matrix.

Input: $\mathbf{X}$, and λ and α, initialization: $\mathbf{Z}^0 = 0$, iteration $i = 0$
Output: representation matrix $\mathbf{Z}^*$ after convergence
while not convergent **do**
  $\mathbf{U}^{i+1} \leftarrow \mathbf{X} - \mathbf{XZ}^i$
  $V^{i+1} \leftarrow \alpha \tanh(\alpha \|\mathbf{U}^{i+1}\|_F^2)$
  $\mathbf{Z}^{i+1} \leftarrow V^{i+1}(V^{i+1}\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^T\mathbf{X}$
end while

C. Proposed Method

1) Learning of the Representation Matrix: In this work, our focus is on the design of robust loss function. For simplicity, we use the Frobenius norm for noise-induced matrix and representation matrix. Based on LHCF, we define:

$$\min_{\mathbf{Z},\mathbf{E}} \ln(\cosh \alpha \|\mathbf{E}\|_F^2) + \lambda \|\mathbf{Z}\|_F^2 \quad \text{s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \tag{4}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and the parameter λ is used to balance the representation matrix term. It is easy to equivalently formulate (4) as

$$\min_{\mathbf{Z}} L = \ln(\cosh \alpha \|\mathbf{X} - \mathbf{XZ}\|_F^2) + \lambda \|\mathbf{Z}\|_F^2. \tag{5}$$

Upon taking the derivative of the above function with respect to the representation matrix, we obtain

$$\frac{\partial L}{\partial \mathbf{Z}} = -2\alpha \tanh(\alpha \|\mathbf{X} - \mathbf{XZ}\|_F^2)\mathbf{X}^T(\mathbf{X} - \mathbf{XZ}) + 2\lambda \mathbf{Z}. \tag{6}$$

By setting (6) to zero, we have

$$(\alpha \tanh(\alpha \|\mathbf{X} - \mathbf{XZ}\|_F^2)\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I})\mathbf{Z} = \alpha \tanh(\alpha \|\mathbf{X}^F - \mathbf{XZ}\|_F^2)\mathbf{X}^T\mathbf{X}. \tag{7}$$

Then, we obtain

$$\begin{cases} \mathbf{U} = \mathbf{X} - \mathbf{XZ} \\ V = \alpha \tanh(\alpha \|\mathbf{U}\|_F^2) \\ \mathbf{Z} = V(V\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^T\mathbf{X}, \end{cases} \tag{8}$$

where $\mathbf{U}$ is the residual matrix, and V is served to lessen the negative effects of complex noise. We can see that the calculation of V relies on the information of representation matrix $\mathbf{Z}$, so it is natural to update $\mathbf{Z}$ in an iterative way. The update process of (8) is summarized in Algorithm 1. It is readily apparent that the computational complexity of learning the representation matrix at each iteration is $O(n^3)$.

2) Subspace Clustering by LHCF: Given that the representation matrix $\mathbf{Z}^*$ is obtained in Algorithm 1, we next utilize the spectral clustering framework to find clustering results. First, the similarity matrix $\mathbf{W}$ is calculated by $\mathbf{W} = (|\mathbf{Z}^*| + |\mathbf{Z}^*|^T)/2$, where $(\cdot)^T$ serves to transpose the matrix. Then, via Normalized Cuts [29], we are able to complete the clustering task based on $\mathbf{Z}$. For more details, one can refer to Algorithm 2.

III. Theoretical Analysis

In order to illustrate that our method is able to correctly find clusters, we theoretically analyze the grouping effect as well as

Algorithm 2: Subspace Clustering using LHCF.

Input: data matrix $\mathbf{X}$, number of clusters k
Output: data clustering results
1: Obtain $\mathbf{Z}^*$ by Algorithm 1.
2: Compute $\mathbf{W}$ by $\mathbf{W} = (|\mathbf{Z}^*| + |\mathbf{Z}^*|^T)/2$.
3: Apply Ncut to $\mathbf{W}$ to accomplish the clustering task.

the convergence behavior in this section. Before performing the analysis, it is necessary to reformulate (5) as

$$\min_{\mathbf{z}_1,\ldots,\mathbf{z}_n} L(\mathbf{Z}) = \ln\left(\cosh\alpha\sum_{j=1}^{n}\|\mathbf{x}_j - \mathbf{X}\mathbf{z}_j\|_2^2\right) + \lambda\sum_{j=1}^{n}\|\mathbf{z}_j\|_2^2, \tag{9}$$

where $\mathbf{z}_j$ is the representation vector of $\mathbf{x}_j$.

A. Grouping Effect

Theorem 1: Given a data vector $\mathbf{x} \in \mathbb{R}^d$, the normalized data matrix $\mathbf{X} \in \mathbb{R}^{d\times n}$, parameters α and λ. Let $\mathbf{z}^*$ be the optimal solution to the following problem :

$$\min_{\mathbf{z}} \ln(\cosh\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}\|_2^2) + \lambda\|\mathbf{z}\|_2^2. \tag{10}$$

Then, one obtain

$$\frac{|z_p^* - z_q^*|}{\|\mathbf{x}\|_2} \le \frac{\alpha}{\lambda}\sqrt{2(1-r)}, \tag{11}$$

where z_p^* and z_q^* are the pth and qth elements of $\mathbf{z}^*$, respectively, and $r = \mathbf{x}_p^T\mathbf{x}_q$, with $\mathbf{x}_p$ and $\mathbf{x}_q$ being the pth and qth columns of $\mathbf{X}$, respectively.

Proof: Let

$$\phi(\mathbf{z}) = \ln(\cosh\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}\|_2^2) + \lambda\|\mathbf{z}\|_2^2. \tag{12}$$

Due to $\mathbf{z}^* = \arg\min_{\mathbf{z}} \phi(\mathbf{z})$, we have

$$\left.\frac{\partial\phi(\mathbf{z})}{\partial\mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}^*} = 0. \tag{13}$$

such that

$$\begin{cases}-2\alpha\tanh(\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2)\mathbf{x}_p^T(\mathbf{x} - \mathbf{X}\mathbf{z}^*) + 2\lambda z_p^* = 0\\ -2\alpha\tanh(\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2)\mathbf{x}_q^T(\mathbf{x} - \mathbf{X}\mathbf{z}^*) + 2\lambda z_q^* = 0.\end{cases} \tag{14}$$

Subtracting the above two equations from each other, we arrive at

$$z_p^* - z_q^* = \frac{\alpha\tanh(\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2)\left[(\mathbf{x}_p^T - \mathbf{x}_q^T)(\mathbf{x} - \mathbf{X}\mathbf{z}^*)\right]}{\lambda}. \tag{15}$$

Taking the absolute value of (15) gives rise to

$$\begin{aligned}|z_p^* - z_q^*| &= \left|\frac{\alpha\tanh(\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2)}{\lambda}\right| \left|(\mathbf{x}_p^T - \mathbf{x}_q^T)(\mathbf{x} - \mathbf{X}\mathbf{z}^*)\right|\\ &\le \frac{\alpha}{\lambda}\left|(\mathbf{x}_p^T - \mathbf{x}_q^T)(\mathbf{x} - \mathbf{X}\mathbf{z}^*)\right|.\end{aligned} \tag{16}$$

It should be aware that the resulted inequality is inspired by $0 \le |\tanh(\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2)| \le 1$. Then, according to the Cauchy-Schwarz inequality, we further have

$$|z_p^* - z_q^*| \le \frac{\alpha}{\lambda}\|\mathbf{x}_p^T - \mathbf{x}_q^T\|_2\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2. \tag{17}$$

Since each column of $\mathbf{X}$ is normalized, it is straightforward that $\|\mathbf{x}_p^T - \mathbf{x}_q^T\|_2 = \sqrt{2(1-r)}$. As $\mathbf{z}^*$ is the optimal solution to the problem (10), it follows that

$$\begin{aligned}\ln(\cosh\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2) &\le \ln(\cosh\alpha\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2) + \lambda\|\mathbf{z}^*\|_2^2\\ &= \phi(\mathbf{z}^*) \le \phi(\mathbf{0}) = \ln(\cosh\alpha\|\mathbf{x}\|_2^2).\end{aligned} \tag{18}$$

Therefore, $\|\mathbf{x} - \mathbf{X}\mathbf{z}^*\|_2^2 \le \|\mathbf{x}\|_2^2$ holds. Combining (17) and (18), we have

$$|z_p^* - z_q^*| \le \frac{\alpha}{\lambda}\sqrt{2(1-r)}\|\mathbf{x}\|_2. \tag{19}$$

If $\mathbf{x}_p$ and $\mathbf{x}_q$ are highly correlated (r is close to 1), then $|z_p^* - z_q^*|$ is approximately zero, which implies that $\mathbf{x}_p$ and $\mathbf{x}_q$ can be grouped into the same cluster. The proof of grouping effect is completed.

B. Convergence Analysis

In this subsection, we perform the convergence analysis of Algorithm 1 by using the Weiszfeld's method [30] which enables global approximation of a function using a sequence of quadratic function [31]. With (9), the solution $\mathbf{Z}$ in (8) can be equivalently transformed to

$$\mathbf{z}_j = V(V\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{x}_j,\ j = 1, 2, \ldots, n. \tag{20}$$

Now assuming a representation matrix $\mathbf{Z}^i$ at iteration i, we denote $\varphi(\mathbf{z}_j; \mathbf{z}_j^i)$ as the upper bound of $L(\mathbf{z}_j)$, where $L(\mathbf{z}_j)$ is obtained by maintaining other variables in $L(\mathbf{Z})$ fixed. It is required for $\varphi(\mathbf{z}_j; \mathbf{z}_j^i)$ to satisfy the following conditions:

$$\begin{cases}\varphi(\mathbf{z}_j^i; \mathbf{z}_j^i) = L(\mathbf{z}_j^i)\\ \varphi'(\mathbf{z}_j^i; \mathbf{z}_j^i) = L'(\mathbf{z}_j^i).\end{cases} \tag{21}$$

Then, $\varphi(\mathbf{z}_j; \mathbf{z}_j^i)$ takes the form

$$\begin{aligned}\varphi(\mathbf{z}_j; \mathbf{z}_j^i) &= L(\mathbf{z}_j^i) + (\mathbf{z}_j - \mathbf{z}_j^i)^T L'(\mathbf{z}_j^i)\\ &\quad + (\mathbf{z}_j - \mathbf{z}_j^i)^T\psi(\mathbf{z}_j^i)(\mathbf{z}_j - \mathbf{z}_j^i),\end{aligned} \tag{22}$$

with

$$\psi(\mathbf{z}_j^i) = \alpha\tanh(\alpha\|\mathbf{X} - \mathbf{X}\mathbf{Z}^i\|_F^2)\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}, \tag{23}$$

being a symmetric matrix. The Algorithm 1 is guaranteed to be convergent according to Theorem 2 below.

Theorem 2: The function (9) is monotonically decreasing under algorithm 1, i.e., $L(\mathbf{Z}^{i+1}) \le L(\mathbf{Z}^i)$.

Proof: Suppose that $\varphi(\mathbf{z}_j; \mathbf{z}_j^i)$ is a locally convex function with respect to $\mathbf{z}_j$ and exists a minimum value at $\mathbf{z}_j^{i+1}$, then we have

$$\varphi'(\mathbf{z}_j^{i+1}; \mathbf{z}_j^i) = L'(\mathbf{z}_j^i) + 2\psi(\mathbf{z}_j^i)(\mathbf{z}_j^{i+1} - \mathbf{z}_j^i) = 0. \tag{24}$$

By choosing $\mathbf{z}_j^i$ near $\mathbf{z}_j$, we have $L(\mathbf{z}_j) \le \varphi(\mathbf{z}_j; \mathbf{z}_j^i)$, then

$$\begin{aligned}L(\mathbf{z}_j^{i+1}) &\le \varphi(\mathbf{z}_j^{i+1}; \mathbf{z}_j^i)\\ &= L(\mathbf{z}_j^i) + (\mathbf{z}_j^{i+1} - \mathbf{z}_j^i)^T L'(\mathbf{z}_j^i)\\ &\quad + (\mathbf{z}_j^{i+1} - \mathbf{z}_j^i)^T\psi(\mathbf{z}_j^i)(\mathbf{z}_j^{i+1} - \mathbf{z}_j^i).\end{aligned} \tag{25}$$

Fig. 2. Examples of Extended Yale B.

From (24) and (25), we deduce

$$L(\mathbf{z}_j^{i+1}) - L(\mathbf{z}_j^i) \leq -(\mathbf{z}_j^{i+1} - \mathbf{z}_j^i)^T \psi(\mathbf{z}_j^i)(\mathbf{z}_j^{i+1} - \mathbf{z}_j^i) \leq -\lambda \|\mathbf{z}_j^{i+1} - \mathbf{z}_j^i\|_2^2 \leq 0, \tag{26}$$

which implies

$$L(\mathbf{z}_j^{i+1}) - L(\mathbf{z}_j^i) \leq 0. \tag{27}$$

Based on (27) and (9), it is easy to verify

$$L(\mathbf{Z}^{i+1}) \leq L(\mathbf{Z}^i). \tag{28}$$

The proof of convergence analysis is completed.

IV. Experiments

In what follows, we test the new method by conducting experiments on the Extended Yale B dataset. It is challenging to perform clustering on Extended Yale B since this dataset is significantly corrupted by "shadows". The Extended Yale B dataset contains 2414 face images of size 192×168 from 38 subjects. We show some examples of Extended Yale B in Fig. 2, and obviously see that this dataset is noisy. Therefore, it is appropriate to choose this dataset to evaluate the algorithm robustness. We compare our algorithm with some popular algorithms, including the K-Means [32], SSC [20], LRR [21], LSR [23], CASS [25], and RSS [26]. To measure the algorithm performance, the clustering accuracy (CA) and normalized mutual information (NMI) are used as evaluation metrics. In our experiment, we consider the first 3, 4, 5 subjects for subspace clustering tasks. Experiments are conducted on a remote computing platform integrated with Intel(R) Xeon(R) Gold 6240 CPU of 2.60Ghz.

A. Data Preprocessing and Algorithmic Setting

To facilitate the model training, we utilize a typical preprocessing mechanism for raw data. Specifically, we first resize the original images to 32×32, and then normalize the image pixel values to $[0, 1]$. Finally, principal component analysis is used to maintain nearly 98% energy by reducing the data dimension.

For all algorithms involved, their parameters are set to achieve the best performance or the recommended parameters are used in the corresponding papers. For a fair comparison, the associated regularization parameters of all the methods involved (except the suggested parameter setting for RSS in [26]) are carefully tuned over the range $[10^{-5}, 10^5]$ with a grid size $\{10^{-5}, 5 \times 10^{-5}, 10^{-4}, 5 \times 10^{-4}, \ldots, 5 \times 10^4, 10^5\}$. Apart from the regularization parameter in our method, the scaling factor α is also carefully tuned over the range of $[10^{-5}, 10^5]$. The optimal value of α is determined to achieve the best performance.

TABLE I
Clustering Accuracy (%) of Different Subspace Clustering Algorithms

Algorithm	K-Means	SSC	LRR	LSR	CASS	RSS	OURS
3 subjects							
Mean	34.80	93.45	92.27	***95.33***	90.62	92.80	**95.78**
Max	35.03	93.98	92.63	***95.89***	90.62	93.07	**96.09**
4 subjects							
Mean	27.85	92.16	90.25	***94.09***	87.92	93.35	**94.88**
Max	28.22	93.05	91.45	***95.94***	88.98	93.75	**95.98**
5 subjects							
Mean	24.84	85.99	88.12	88.84	81.64	**90.52**	***90.39***
Max	25.45	87.56	89.33	91.22	83.91	**91.67**	***91.56***

TABLE II
Normalized Mutual Information (%) of Different Subspace Clustering Algorithms

Algorithm	K-Means	SSC	LRR	LSR	CASS	RSS	OURS
3 subjects							
Mean	0.10	79.33	76.42	***83.44***	74.27	79.22	**83.84**
Max	0.12	80.01	76.98	***84.27***	74.27	79.63	**85.10**
4 subjects							
Mean	0.94	81.99	79.20	***86.20***	75.34	83.93	**86.76**
Max	1.62	82.80	80.35	***87.65***	76.08	84.19	**87.67**
5 subjects							
Mean	2.65	75.67	78.82	79.93	67.59	**81.94**	***80.56***
Max	4.04	76.75	79.50	81.26	68.88	**82.88**	***81.31***

B. Experimental Results

Tables I and II record the experimental results of CA and NMI of different subspace clustering algorithms, respectively. To highlight some important results, we use regular bold font to mark the best performance, and italic bold font to mark the second best performance.

We can see from Table I that K-means shows the poorest clustering performance. In comparison, SSC and LRR have significant accuracy improvement, but their performance is inferior to LSR, RSS and our algorithm. Our algorithm achieves the best performance on the average accuracy and maximum accuracy in the case of 3 subjects and 4 subjects. Despite that the clustering accuracy of RSS is slightly higher than our algorithm in the case of 5 subjects, the newly proposed algorithm is still a better clustering scheme by comprehensively considering all the cases. In Table II, it is also clear that our algorithm is a better choice for subspace clustering in terms of NMI. Overall, our proposed method is able to provide good clustering results.

V. Conclusion

In this letter, we proposed a novel robust subspace clustering algorithm. Specifically, our method was derived based on the LHCF loss function that is useful in suppressing complex noise. We also analyzed the grouping effect as well as the convergence, theoretically illustrating that highly correlated samples can be grouped into the same cluster. Experimental results conducted on the Extended Yale B dataset have demonstrated the superiority of algorithm.

References

[1] C.-G. Li and R. Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 277–286.

[2] C.-G. Li, C. You, and R. Vidal, "Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2988–3001, Jun. 2017.

[3] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[4] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, Oct. 2010.

[5] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[6] X. Peng, J. Feng, J. T. Zhou, Y. Lei, and S. Yan, "Deep subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5509–5521, Dec. 2020.

[7] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.

[8] F. Wu, P. Yuan, G. Shi, X. Li, W. Dong, and J. Wu, "Robust subspace clustering network with dual-domain regularization," *Pattern Recognit. Lett.*, vol. 149, pp. 44–50, 2021.

[9] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. I–I.

[10] T. Zhang, A. Szlam, and G. Lerman, "Median k-flats for hybrid linear modeling with many outliers," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops*, 2009, pp. 234–241.

[11] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.

[12] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, pp. 413–458, 2008.

[13] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, 1999.

[14] A. Gruber and Y. Weiss, "Multibody factorization with uncertainty and missing data using the EM algorithm," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. I–I.

[15] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 317–330, 2009.

[16] P. Favaro, R. Vidal, and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1801–1807.

[17] Z. Kang, C. Peng, and Q. Cheng, "Robust subspace clustering via smoothed rank approximation," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2088–2092, Nov. 2015.

[18] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, "Subspace clustering by block diagonal representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 487–501, Feb. 2019.

[19] J. Wang, X. Wang, F. Tian, C. H. Liu, and H. Yu, "Constrained low-rank representation for robust subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4534–4546, Dec. 2017.

[20] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[21] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[22] Y. Qin, X. Zhang, L. Shen, and G. Feng, "Maximum block energy guided robust subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2652–2659, Feb. 2023.

[23] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 347–360.

[24] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear-norm and frobenius-norm-based representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.

[25] C. Lu, J. Feng, Z. Lin, and S. Yan, "Correlation adaptive subspace segmentation by trace lasso," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1345–1352.

[26] X. Guo, "Robust subspace segmentation by simultaneously learning data representations and their affinity matrix," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 3547–3553.

[27] X. Li, Q. Lu, Y. Dong, and D. Tao, "Robust subspace clustering by cauchy loss function," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2067–2078, Jul. 2019.

[28] S. Wang, W. Wang, K. Xiong, H. H. Iu, and K. T. Chi, "Logarithmic hyperbolic cosine adaptive filter and its performance analysis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 4, pp. 2512–2524, Apr. 2021.

[29] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[30] H. Voß and U. Eckhardt, "Linear convergence of generalized Weiszfeld's method," *Computing*, vol. 25, no. 3, pp. 243–251, 1980.

[31] I. Singer, *Duality for Nonconvex Approximation and Optimization*. Berlin, Germany: Springer, 2007.

[32] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, 2003.