SUBSPACE CLUSTERING USING UNSUPERVISED DATA AUGMENTATION

Maryam Abdolali *Nicolas Gillis*

Department of Mathematics and Operational Research,
Université de Mons, Belgium

ABSTRACT

Subspace clustering is an unsupervised approach for determining the union of multiple subspaces that best fits a collection of high-dimensional samples. Self-expressive representation, that is, expressing samples as linear combination of other samples, is the core of most state-of-the-art subspace clustering approaches. However, existence of sufficiently well-spread samples within each subspace is crucial for precise representation, which might not always be available in real-world scenarios. Inspired by the remarkable influence of data augmentation on the performance of neural networks, we propose a scalable approach that employs data augmentation within subspace clustering. Benefiting from the increased diversity in data, we use augmented samples as an enlarged dictionary and combine the self-expressive representations based on the assumption that augmentation does not alter the labels of the samples. Significant improvement of the clustering performance on two real-world datasets demonstrates the effectiveness of the proposed approach.

***Index Terms*—** subspace clustering, data augmentation, unsupervised learning, sparse representation, clustering

1. INTRODUCTION

Data clustering is one of the fundamental tasks in unsupervised machine learning. Among the many clustering methodologies, subspace clustering has been established as a major approach for grouping samples according to their underlying subspaces. Subspace clustering is based on the assumption that the samples approximately lie on several low-dimensional subspaces, and is defined as follows [1]:

Definition 1 (Subspace clustering) *Let $X \in \mathbb{R}^{d\times n}$ be the data matrix containing n samples of dimension d, where each sample, $X(:,i)$ $(i = 1, ..., n)$, is a column of this matrix. Assume that the samples are distributed on the union of unknown r linear subspaces, $\{S_i\}_{i=1}^r$, with unknown dimensions $\{d_i\}_{i=1}^r$, with $d_i < d$. The goal of subspace clustering is to partition the samples according to their underlying subspaces.*

The authors acknowledge the support by the ERC (StG no 679515), and by the F.R.S.-FNRS and the FWO under EOS Project no O005318F-RG47.

In the past two decades, a wide variety of algorithms have been proposed to tackle subspace clustering; see, e.g., [2, 3, 4, 5]. Among them, the family of approaches based on exploiting the self-expressive property is the current state-of-the-art [5, 6, 7]. Self-expressiveness assumes that every sample on a (linear) subspace/cluster can be represented as a linear combination of other samples in the same subspace. These approaches require solving an optimization problem of the form:

$$\min_C \ \theta(C) + \lambda \, \mathcal{L}(X - XC) \ \text{ such that } \ \operatorname{diag}(C) = 0, \quad (1)$$

where $C \in \mathbb{R}^{n\times n}$ is the coefficient matrix, $\theta(C)$ is a regularizer (e.g., to promote sparsity), $\lambda > 0$ is a regularization parameter, and $\mathcal{L}$ is the data fitting term, which is typically the squared Frobenius norm $\mathcal{L}(X-XC) = \|X-XC\|_F^2$. The diagonal entries of the matrix C, that is, $\operatorname{diag}(C)$, are set to zero in order to avoid the trivial solution (the identity matrix). Several different regularizers have been proposed in the literature, the three most widely used are: (1) the component-wise ℓ_1 norm, $\|C\|_1$, in sparse subspace clustering (SSC) [5], (2) the nuclear norm, $\|C\|_*$, in low-rank representation (LRR) [6], and (3) the Frobenius norm, $\|C\|_F^2$, in least square regression (LSR) [7]. After having solved (1), each entry of the coefficient matrix, $C(i,j)$ for $i,j \in \{1, ..., n\}$, provides an estimation of pairwise similarity between the two corresponding samples, $X(:,i)$ and $X(:,j)$. Finally, spectral clustering is applied on the symmetric affinity matrix corresponding to C, typically constructed as $|C| + |C^\top|$, to obtain the final segmentation of the samples, where $|\cdot|$ is the element-wise absolute value.

Despite the notable performances on several real-world datasets, self-expressive representations heavily rely on the existence of sufficiently well-spread samples from each subspace to obtain a *subspace preserving* representation. A coefficient matrix is subspace preserving if each sample only uses the samples from the same subspace [5]. It has been shown that the data distribution within each subspace can significantly affect the performance of subspace clustering algorithms [5]. However, to the best of our knowledge, most subspace clustering approaches consider the given samples as a fixed input, and focus on improving the quality of the coefficient matrix given the samples.

Data augmentation has been shown to be extremely effective in supervised and lately unsupervised learning in neural networks [8, 9, 10]. In data augmentation, new data is synthetically produced from the given data. Typically, the new data is obtained by some transformation of the given samples in such a way that the category/cluster of the samples remain unchanged. Inspired by their remarkable success in boosting the performance of many neural-network based approaches, we combine the advantages of (unsupervised) data augmentation with subspace clustering. The goal is that by leveraging invariances, additional generated samples can help to avoid degenerate subspaces and lead to well-spread samples.

Incorporating data augmentation techniques with subspace clustering is mostly overlooked in the literature. To the best of our knowledge, [11] is the only existing approach that combines the benefits of augmentation with the previously proposed deep subspace clustering framework [12]. In particular, the representations are learned such that they lead to consistent subspaces for different augmentation strategies. However, pursuing subspace consistency among different sets of augmented samples can be interpreted as a *multi-view* subspace clustering where each set of newly generated data is a different view of the original samples. Instead of focusing separately on different sets of augmented samples, the main goal of this paper is to propose an approach that computes the coefficient matrix using the whole augmented data as an enlarged dictionary to represent the given samples.

This paper is organized as follows. In Section 2, we present the detailed description of the proposed approach, dubbed augmented k-nearest neighbors subspace clustering (AK-SC), and provide a geometric illustration on how augmentation can improve the quality of the coefficient matrix, especially as the affinity between subspaces increases. In Section 3, we provide numerical experiments illustrating the effectiveness of AK-SC compared to state-of-the-art subspace clustering algorithms.

2. AUGMENTED SELF-EXPRESSIVE REPRESENTATION

Generally, the goal of data augmentation is to increase the data diversity using the existing samples. How to generate new artificial samples from the given data, *without alternating the category* of the corresponding sample, is an active research direction. In this paper, we focus on image data and utilize classic augmentation transformations (such as flip, rotation and scaling). In this section, we first present the proposed approach for merging the benefits of augmentation with subspace clustering in a unified optimization model. Then the geometrical interpretation for augmented self-expressive representation is provided.

2.1. Augmented k-nearest neighbors subspace clustering

Let $\hat{X} = [\hat{X}_1, ..., \hat{X}_m] \in \mathbb{R}^{d\times nm}$ be the generated augmented data by applying m different augmentation strategies on the input data matrix X, where $\hat{X}_i \in \mathbb{R}^{d\times n}$ is the generated data by the i-th strategy, for $i = 1, ..., m$. We form the final overcomplete dictionary $\tilde{X}$ by concatenating the original data, X, and the augmented samples, $\hat{X}$, as $\tilde{X} = [X,\ \hat{X}]$. Using the new extracted information in $\tilde{X}$, our proposed sparse self-expressive representation is formulated as follows:

$$\begin{aligned}\tilde{C} \quad = \quad & \underset{C\in\mathbb{R}^{n(m+1)\times n}}{\operatorname{argmin}} \|C\|_1 + \frac{\lambda}{2}\|X - \tilde{X}C\|_F^2 \\ & \text{such that } C(j, \Omega(j)) = 0 \text{ for } j = 1, ..., n,\end{aligned} \tag{2}$$

where $\Omega(j)=\{j+(k-1)n\}_{k=1}^{m+1}$ contains j and the indices of the augmented samples corresponding to $X(:, j)$. By setting the corresponding entries in $\tilde{C}$ to zero, we avoid representing a sample using its augmented samples. On the other hand, the rectangular matrix $\tilde{C}$ has a specific block-wise structure, that is, $\tilde{C} = [\tilde{C}_1\ ;\ ...\ ;\ \tilde{C}_{m+1}]$ which is obtained by vertical concatenation of sub-matrices $\{\tilde{C}_i\}_{i=1}^{m+1} \in \mathbb{R}^{n\times n}$. In particular, $\tilde{C}_1$ contains the representation coefficients between the original samples in X, and $\{\tilde{C}_i\}_{i=2}^{m+1}$ between the original samples and the augmented samples, $\{\hat{X}_i\}_{i=1}^{m}$. We construct the final coefficient matrix as follows $C_f = \sum_{i=1}^{m+1} \tilde{C}_i$, and the cluster assignments are obtained by applying spectral clustering on the affinity matrix $A_f = |C_f| + |C_f^\top|$. The matrix C_f merges the representation coefficients using the original data and all augmented samples in a simple unsupervised way. This is based on the assumption that the augmented samples should belong to the same clusters of the corresponding original data, and representing the data using the augmented samples indicates a connection/link to the corresponding original sample. As we will see, this simple strategy performs rather efficiently. Designing more sophisticated aggregation strategies is a topic of further research.

Augmentation leads to an enlarged dictionary, $\tilde{X}$, which increases the computational complexity. To reduce the computational burden of dealing with the augmented samples, similar to [20], we impose a sparsity structure on $\tilde{C}$, namely each sample is only allowed to use columns of the dictionary corresponding to the k nearest neighbors. Let $\mathcal{N}_k(X(:, j))$ be the k nearest neighbors of $X(:, j)$ in $\tilde{X}$ and Ψ_j be the set containing the indices of these samples. We ensure that $\Psi_j \cap \{j + (k-1)n\}_{k=1}^{m+1} = \emptyset$, that is, no augmented sample of $X(:, j)$ is included in $\mathcal{N}_k(X(:, j))$. Finally, for each $j = 1, 2, \ldots, n$, we solve the following problem:

$$\bar{C}(:, j) \ = \ \underset{c\in\mathbb{R}^k}{\operatorname{argmin}} \|c\|_1 + \frac{\lambda}{2}\|X(:, j) - \mathcal{N}_k(X(:, j))c\|_2^2. \tag{3}$$

An additional advantage of limiting the dictionary to the k nearest neighbors is that the samples are represented using *spatially close* samples which promotes locality preserving

representations [13]. The sparse $n(m+1)\times n$ matrix $\tilde{C}$ is constructed by setting $\tilde{C}(\Psi_j, j) = \bar{C}(:, j)$ for $j = 1, ..., n$. We use an ADMM based algorithm [14] to solve (3); see Algorithm 1. We refer to our proposed algorithm as augmented k-nearest neighbors subspace clustering (AK-SC): AK-SC first solves (3) using Algorithm 1 to obtain C_f, and then performs spectral clustering on $|C_f| + |C_f^\top|$.

Algorithm 1 ADMM for (3) in AK-SC

Input: $X \in \mathbb{R}^{d\times n}$, m predefined augmentation strategies, parameters λ and k.

Output: Pairwise coefficient matrix C_f.

Initialization: Apply m augmentation strategies to obtain $\tilde{X} \in \mathbb{R}^{d\times nm}$, $\bar{C} = A = \Delta = 0$, $\mu = \frac{\lambda}{\max_{i\neq j} |x_j^T x_i|}$, $\rho = \lambda$
for each sample $X(:, j)$ in X **do**
  Set $D = \mathcal{N}_k(X(:, j))$ as the dictionary, Ψ_j as the indices of $\mathcal{N}_k(X(:, j))$ with $\Psi_j \cap \{j+(k-1)n\}_{k=1}^{m+1} = \emptyset$.
  while some convergence criterion is not met **do**
    $A(:, j) \leftarrow \left(\mu D^\top X(:, j) + \rho I_k\right)^{-1} (\mu D^\top X(:, j) + \rho\bar{C}(:, j) - \Delta(:, j))$ where I_k is the identity matrix of dimension k.
    $\bar{C}(:, j) \leftarrow T_{\frac{1}{\rho}}(A(:, j) + \Delta(:, j)/\rho)$ where $T_\gamma(y) = \max(0, |y| - \gamma)\,\mathrm{sign}(y)$ is the soft-thresholding operator.
    $\Delta(:, j) = \Delta(:, j) + \rho(A(:, j) - \bar{C}(:, j))$
  end while
  Set $\tilde{C}(\Psi_j, j) = \bar{C}(:, j)$
end for
Set C_f as block-wise sum of submatrices in $\tilde{C}$.

2.2. Geometrical motivation of self-expressive representation using augmented data

In the noiseless case, solving (1) using the component-wise ℓ_1-norm regularizer for the coefficient matrix has an interesting geometrical interpretation. Let $\mathcal{P}(\pm X_{-j})$ be the symmetrized convex hull of the columns of the matrix X except for the jth sample. It has been proved that the nonzero entries in the coefficient vector, $C(:, j)$, correspond to the vertices of the face of $\mathcal{P}(\pm X_{-j})$ that intersects with the ray generated by $X(:, j)$, that is, the line that goes from the origin to the point $X(:, j)$ [15, 5].

Following this geometric interpretation, we claim that generating new samples using augmentation can improve the accuracy, especially when dealing with close subspaces (high affinity). This is illustrated in Figure 1. Suppose the samples (shown in full circles) are drawn from three disjoint linear subspaces (a plane and two lines). Let x be a point on the intersection of S_1 and $S_2 \bigoplus S_3$. The symmetrized convex hull of the points on S_1 (discarding x) and the points on $S_2 \bigoplus S_3$ are shown in dotted blue and magenta, respectively. The ray passing through x reaches the symmetrized convex hull of $S_2 \bigoplus S_3$ later than the one corresponding to S_1, and hence is represented by the sparse representation of samples from the wrong subspaces. However, after introducing appropriate augmented samples (the crosses on Figure 1), the sample x is represented correctly by the samples from the same subspace. As the affinity between subspaces increases, we have observed that the augmented samples typically become more beneficial in improving the quality of the coefficient matrix.

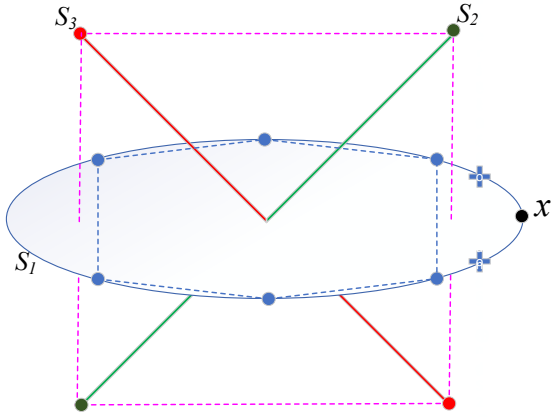

Fig. 1. Illustration of the role of augmented samples in self-expressive based subspace clustering. The original samples are drawn from three disjoint subspaces $\{S_i\}_{i=1}^3$ and shown in full circle. The corresponding symmetrized convex hull of the points on the S_1 (without x) and $S_2 \bigoplus S_3$ are shown in blue and magenta dotted lines, respectively. Using the original samples, the sample x is represented wrongly by samples from $S_2 \bigoplus S_3$. However, by introducing augmented samples, the crosses within S_1, the ray generated by x reaches the face of S_1 later than the one in $S_2 \bigoplus S_3$, and hence is correctly represented by samples from the same subspace.

3. NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of the proposed AK-SC for clustering the images of two real-world datasets: COIL-20 (Columbia Object Image Library) [16] and MNIST. COIL-20 contains 72 processed images of size 128×128 of each of the 20 objects, for a total of 1440 images. The images were taken from objects in different poses. We downsampled them to 32×32 pixels as in [12]. The MNIST dataset contains 70,000 images with 28×28 pixels of 10 handwritten digits. The dataset is divided into 60,000 images for training and 10,000 images for testing. We use 50 images per digit (cluster) from the MNIST test dataset. The performance of AK-SC is compared with the following popular subspace clustering algorithms: SSC [5], EnSC [17], OMP [18], KSSC [19], and knn-SSC [20]. The parameters are selected following the recommendations in the corresponding papers to obtain the best

performance. We did not compare with the deep subspace clustering approach with data augmentation from [11] as the implementation is not available. Moreover, [11] is based on introducing a linear self-expressive based layer in an autoencoder network, which is shown in [21] to be ill-posed and lead to a degenerate embedding of the data. The code to run the experiments is available from `https://bit.ly/AKSC_v1`.

We evaluate the performances by measuring the clustering error. Let $\hat{\ell} \in \{1, \dots, r\}^n$ be the obtained labels (clusters) by a clustering algorithm (recall that r is the number of clusters). Given the ground-truth labels $\ell \in \{1, \dots, r\}^n$, the clustering error is calculated as follows:

$$\text{err} = \min_{\pi, \text{ a permutation}} \frac{\left|\{i \mid \ell(i) \neq \hat{\ell}_\pi(i)\}\right|}{n},$$

where $\hat{\ell}_\pi$ are the labels obtained after permuting using π.

For the MNIST dataset, we follow the common practice of applying a scattering convolution network [22] on the data to extract feature vectors of dimension 3472, and then project them to dimension 100 using PCA [18]. Nine datasets corresponding to different number of digits (from 0 to $1, \dots, 9$) are constructed. The clustering errors of the different algorithms are reported in Table 1. For the proposed AK-SC, we used two standard augmentation strategies: 5 *random rotations* within the range $[-30^\circ, 30^\circ]$, and 5 *random scalings* within the range $[0.8, 1.2]$, and hence a total of 10 augmented samples. Due to the random nature of the augmented strategies, the average and standard deviation of ten trials are reported for AK-SC. We set the parameters of AK-SC as $k = 30$ and $\lambda = 100$. We did not notice major sensitivity to the parameter k. For instance, for $k = 20$ (resp. 40), we obtained an average error rate of 8.00% (resp. 8.49%) for clustering the digits $[0 : 9]$.

Table 1. Clustering error rate (%) on the MNIST dataset (best performance in bold, second best underlined).

Digits	SSC	OMP	EnSC	knn-SSC	KSSC	AK-SC
[0:1]	**0**	**0**	0.1	**0**	**0**	**0±0**
[0:2]	**0**	0.62	**0**	0.67	**0**	2.93±0.56
[0:3]	32.50	4.00	32.50	4.50	**1.50**	2.05±0.43
[0:4]	27.20	30.00	27.60	4.00	**1.60**	2.08±0.41
[0:5]	22.33	23.33	22.33	9.33	7.71	**3.33±0.27**
[0:6]	21.14	22.29	19.43	8.86	7.71	**3.62±0.55**
[0:7]	19.25	21.25	17.50	8.75	7.75	**3.67±0.48**
[0:8]	18.22	23.56	18.67	21.33	14.22	**6.02±0.26**
[0:9]	18.00	23.20	18.20	22.40	15.60	**7.90±0.80**

For the COIL-20 dataset, we use the augmentation strategies of *flipping left to right*, 5 *random rotations* within the range $[-30^\circ, 30^\circ]$, and 5 *random scalings* within the range $[0.8, 1.2]$. We set $k = 10$ and $\lambda = 20$ for AK-SC. We observed that for larger values of k, more augmented samples are needed to keep the error rate unchanged. The error rate of different algorithms are reported in Table 2. The results of AK-SC are also averaged over ten different trials.

Table 2. Clustering error rate (%) on the COIL-20 dataset (best performance in bold, second best underlined).

SSC	OMP	EnSC	knn-SSC	KSSC	AK-SC
24.38	29.86	12.40	20.56	14.94	**0±0**

Looking at Tables 1 and 2, we observe the following:

• AK-SC significantly reduces the error rate for COIL-20, from 12.40% for the best competitor (namely EnSC) to 0% for AK-SC, for the 10 randomly generated augmentations. The reason is that augmentation allows to fill in the subspaces by creating augmented samples to help representing the original images using samples from the same subspace, as explained in Section 2.2. This is particularly important for nearby subspaces: the main struggle on COIL-20 for the majority of (subspace) clustering algorithms is to distinguish between three similar categories of toy cars. A reason AK-SC performs particularly well for COIL-20 is that the original images in COIL-20 are already some kind of rotations/scalings of the different objects (more precisely, they are taken from different angles). In fact, it turns out the graph corresponding to the affinity matrices of AK-SC has no wrong connections between samples, and has sparsely connected samples within each cluster. On the other hand, knn-SSC, which is similar to AK-SC without augmentation, has many wrong connections between the samples of different clusters. This shows that augmentation plays a significant role in boosting the performance, which we also observe on MNIST.

• For the MNIST dataset, as the number of clusters increases, particularly starting from $[0 : 5]$, AK-SC consistently outperforms all other algorithms by a large margin.

• The general considerable decrease in error rate suggests that augmentation offers more than several *multi-view* representations of the original data. In other words, focusing separately on each augmented set and merely seeking *consistency* between the corresponding coefficient matrices neglects the complementary information within the augmented samples.

4. CONCLUSION

In this paper, we proposed AK-SC to integrate augmented samples with sparse subspace clustering. Enlarging the data using augmentation strategies that do not change the subspace of the samples provides additional information that can significantly benefit the self-expressive representation. The augmented data serves as an overcomplete dictionary in AK-SC and the computational cost is controlled by limiting the representation of each sample to a few nearest neighbors. We geometrically illustrated how augmentation can improve the self-expressive representation, which we have illustrated on two widely used datasets. In this paper, we focused on utilizing predefined standard augmentation strategies. An intriguing research direction is to learn the augmentation strategies from the samples without altering the underlying subspaces.

5. REFERENCES

[1] René Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 52–68, 2011.

[2] Paul S Bradley and Olvi L Mangasarian, "K-plane clustering," *Journal of Global Optimization*, vol. 16, no. 1, pp. 23–32, 2000.

[3] Michael E Tipping and Christopher M Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

[4] Ken-ichi Kanatani, "Motion segmentation by subspace separation and model selection," in *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2001, vol. 2, pp. 586–591.

[5] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[6] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2012.

[7] Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan, "Robust and efficient subspace segmentation via least squares regression," in *European Conference on Computer Vision*. Springer, 2012, pp. 347–360.

[8] Agnieszka Mikołajczyk and Michał Grochowski, "Data augmentation for improving deep learning in image classification problem," in *International Interdisciplinary PhD Workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.

[9] Connor Shorten and Taghi M Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[10] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V Le, "Unsupervised data augmentation for consistency training," *arXiv preprint arXiv:1904.12848*, 2019.

[11] Mahdi Abavisani, Alireza Naghizadeh, Dimitris N Metaxas, and Vishal M Patel, "Deep subspace clustering with data augmentation.," in *NeurIPS*, 2020.

[12] Pan Ji, Tong Zhang, Hongdong Li, Mathieu Salzmann, and Ian Reid, "Deep subspace clustering networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 24–33.

[13] Maryam Abdolali and Nicolas Gillis, "Beyond linear subspace clustering: A comparative study of nonlinear manifold clustering algorithms," *arXiv preprint arXiv:2103.10656*, 2021.

[14] Daniel Gabay and Bertrand Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[15] Behrooz Nasihatkon and Richard Hartley, "Graph connectivity in sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2137–2144.

[16] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al., "Columbia object image library (COIL-100)," 1996.

[17] Chong You, Chun-Guang Li, Daniel P Robinson, and René Vidal, "Oracle based active set algorithm for scalable elastic net subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3928–3937.

[18] Chong You, Daniel Robinson, and René Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3918–3927.

[19] Vishal M Patel and René Vidal, "Kernel sparse subspace clustering," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2849–2853.

[20] Liansheng Zhuang, Jingjing Wang, Zhouchen Lin, Allen Y Yang, Yi Ma, and Nenghai Yu, "Locality-preserving low-rank representation for graph construction from nonlinear manifolds," *Neurocomputing*, vol. 175, pp. 715–722, 2016.

[21] Benjamin David Haeffele, Chong You, and Rene Vidal, "A critique of self-expressive deep subspace clustering," in *International Conference on Learning Representations (ICLR)*, 2021.

[22] Joan Bruna and Stéphane Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.