NULL SPACE CLUSTERING WITH APPLICATIONS TO MOTION SEGMENTATION AND FACE CLUSTERING

Pan Ji[1], *Yiran Zhong*[1], *Hongdong Li*[1], *Mathieu Salzmann*[1,2]

[1]Australian National University, Canberra
[2]NICTA, Canberra

{pan.ji,hongdong.li}@anu.edu.au,mathieu.salzmann@nicta.com.au

ABSTRACT

The problems of motion segmentation and face clustering can be addressed in a framework of subspace clustering methods. In this paper, we tackle the more general problem of clustering data points lying in a union of low-dimensional linear(or affine) subspaces, which can be naturally applied in motion segmentation and face clustering. For data points drawn from linear (or affine) subspaces, we propose a novel algorithm called Null Space Clustering (NSC), utilizing the null space of the data matrix to construct the affinity matrix. To better deal with noise and outliers, it is converted to an equivalent problem with Frobenius norm minimization, which can be solved efficiently. We demonstrate that the proposed NSC leads to improved performance in terms of clustering accuracy and efficiency when compared to state-of-the-art algorithms on two well-known datasets, i.e., Hopkins 155 and Extended Yale B.

Index Terms— null space, subspace clustering, affinity matrix, normalized cuts, motion segmentation, face clustering

1. INTRODUCTION

It's well known that under affine camera model, the trajectories of each motion correspond to a four dimensional linear subspace (or a three dimensional affine subspace) [1], and under Lambertian reflectance assumption, faces of the same subject captured with a fixed pose and varying light sources lie approximately in a ten dimensional linear subspace [2]. Thus the problems of motion segmentation and face clustering (See Fig 1) become a problem of subspace clustering. In this paper, we consider the genenral problem of clustering the data points according to their respective subspaces, i.e., finding the membership of the data points to the subspaces. This problem has been proven important, with a lot of applications in image processing [3, 4]. Up to now, many different approaches addressing this problem have been proposed, including factorized-based algorithms [5, 6], algebraic methods [7], and spectral clustering based algorithms [8, 9, 10, 11].

One of the earliest attempts in this research area was the factorization based algorithm [6], which got subspace clustering by a low-rank factorization of the data matrix $\mathbf{X}$. To obtain the membership of each data point, the so-called Shape Interaction Matrix (SIM) was proposed, which is defined as $\mathbf{Q} = \mathbf{V}_1\mathbf{V}_1^T$ with $\mathbf{V}_1$ being the right singular vector (from the reduced Singular Value Decomposition) of $\mathbf{X}$. Then we can obtain the clustering results by simply block-diagonalizing $\mathbf{Q}$ [6]. However, this algorithm cannot handle noise and outliers explicitly, and thus is not robust enough in practice.

Later, algebraic methods, such as Generalized Principal Component Analysis (GPCA) [7] and Robust Algebraic Segmentation (RAS) [12], were proposed, which fit polynomials to data points. Thus subspace clustering problem, in this case, becomes a problem of fitting and differentiating polynomials. Unfortunately, the complexity of these methods will increase exponentially with the number and the dimension of the subspaces.

Recently, there has been a surge of spectral clustering (or normalized cuts) based methods, among which notable methods are Sparse Subspace Clustering (SSC) [1, 8], Low Rank Representation (LRR) [13, 9], Low Rank Subspace Clustering (LRSC) [11], Subspace Segmentation by Least Square Regression (LSR) [14], and Efficient Dense Subspace Clustering (EDSC) [10]. These approaches often consist of two steps: first, build an affinity matrix such that only points in the same subspace are connected; second, apply spectral clustering with this affinity matrix. The key to the success of these approaches resides in building a good affinity matrix, so most recent work focuses on the first step. Typically, the affinity matrix is constructed utilizing data's property of self-expressiveness, i.e., the data matrix can be taken as the dictionary to represent the data itself. In particular, SSC recovers a sparse representation of each data point via an ℓ_1 norm minimization formulation to construct an affinity matrix whose non-zeros entries correspond to points in the same subspace, whereas LRR, LRSC, LSR and EDSC get dense representations by rank minimization (LRR, LRSC) or Frobenius norm minimization (LSR, EDSC).

In this paper, we propose a novel subspace clustering algorithm called Null Space Clustering (NSC), which is also a spectral clustering based algorithm. Our key observation is that the null space of the data matrix contains the desirable information for subspace clustering and the orthogonal projector of the null space $\mathbf{V}_2\mathbf{V}_2^T$ forms block-diagonal structure, i.e., only points on the same subspaces will be connected. To better deal with noise and outliers, it is converted to an equivalent formulation with Frobenius norm minimization, which allows us to model the noise and outliers explicitly and can also be solved efficiently. We demonstrate the effectiveness and efficiency of our algorithm on motion segmentation and face clustering. In particular, we get state-of-the-art results on two well known datasets, i.e., Hopkins 155 motion segmentation dataset and the Extended Yale B face dataset [15].

2. BACKGROUND

In this section, we introduce the background of subspace clustering problem and explain the related algorithms, e.g., SSC [8], LRR [9] and LSR [14].In the following part of the paper, $\mathbf{X} \in \mathbb{R}^{D\times N}$ denotes the data matrix whose columns are the data points drawn from a union of k independent subspaces and assume columns of $\mathbf{X}$ have been sorted according to the subspaces.

2.1. ℓ_p Norm Minimization Based Algorithms

The algorithms with ℓ_p norm minimization can be compactly formulated as

$$\min_{\mathbf{C}} \|\mathbf{C}\|_p \quad \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{C} \; (\text{diag}(\mathbf{C}) = 0) \; . \tag{1}$$

Then SSC, LRR and LSR become the special cases of Eq. (1). When $p = 1$, Eq. (1) turns out to be the noise free formulation for SSC, which represents each data point as a sparse linear combination of all other points. In this case, the constraint $\text{diag}(\mathbf{C}) = 0$ should be enforced to avoid trivial solution. When p denotes nuclear norm ($\|\cdot\|_*$) or Frobenius norm ($\|\cdot\|_F$), it results in the noise free formulations for LRR and LSR respectively, which have been shown to be two equivalent formulations sharing the same solution $\mathbf{C}^* = \mathbf{V}_1\mathbf{V}_1^T$, c.f. [10]. This solution again coincides with SIM, c.f. [10]. In these two cases of LRR and LSR, the diagonal constraint $\text{diag}(\mathbf{C}) = 0$ is not necessary since it obviously will not lead to trivial solution even without this constraint. Then the affinity matrix can be built as $\mathbf{A} = |\mathbf{C}| + |\mathbf{C}^T|$, and the clustering result can be obtained by applying normalized cuts algorithm [16] on $\mathbf{A}$.

2.2. Null Space Formulation

We now introduce our null space formulation. Let $\mathbf{V}_2$ be the null space of data matrix $\mathbf{X}$. In general, each column of $\mathbf{V}_2$ is a linear combination of the basis of the subspaces, and thus $\mathbf{V}_2$ does not exhibit a block-diagonal structure. However, the orthogonal projector constructed by $\mathbf{V}_2$ preserves the structure of the original subspaces, as defined below:

$$\mathbf{W} = \mathbf{V}_2\mathbf{V}_2^T \; . \tag{2}$$

The matrix $\mathbf{W}$ can be proved to be block-diagonal in a similar way to SIM [6]. Then, the affinity matrix can be constructed as $\mathbf{A} = |\mathbf{W}| + |\mathbf{W}^T|$, or via some intuitive methods to enhance the block structure of $\mathbf{C}$ as in [17, 13].

3. PROBLEM REFORMULATION WITH FROBENIUS NORM MINIMIZATION

The key drawback of formulation (2) is that the noise and outlier cannot be explicitly handled, thus it is sensitive to them. When the data is contaminated by noise or outliers, it becomes hard to recover the true null space of the clean data matrix. However, we can show that the null space clustering problem can be equivalently reformulated as:

$$\min_{\mathbf{C}} \|\mathbf{I} - \mathbf{C}\|_F^2 \quad \text{s.t.} \quad \mathbf{X}\mathbf{C} = \mathbf{0}, \tag{3}$$

which makes it easy to handle noise and outliers explicitly, as will be shown in the following subsections.

We then show by a lemma that formulations (2) and (3) are equivalent.

Lemma 1. *The solution to* (3) *is:* $\mathbf{C}^* = \mathbf{V}_2\mathbf{V}_2^T$*, which is equal to* $\mathbf{W}$ *defined in* (2)*.*

Proof. The linear constraints in (3) can be eliminated by expressing $\mathbf{C} = \mathbf{V}_2\Phi$, where Φ is the new variable. Then

$$\Phi^* = \underset{\Phi}{\text{argmin}} \, \|\mathbf{I} - \mathbf{V}_2\Phi\|_F^2 \tag{4}$$

$$= \underset{\Phi}{\text{argmin}} \, \text{trace}(-2\mathbf{V}_2\Phi + \Phi^T\mathbf{V}_2^T\mathbf{V}_2\Phi) \tag{5}$$

$$= \underset{\Phi}{\text{argmin}} \, \text{trace}(-2\mathbf{V}_2\Phi + \Phi^T\Phi) \tag{6}$$

Take the derivative over Φ and set it to zero, yielding $\Phi^* = \mathbf{V}_2^T$. So $\mathbf{C} = \mathbf{V}_2\Phi^* = \mathbf{V}_2\mathbf{V}_2^T$, which concludes the proof. □

In the following subsections, we'll show how to handle noise, outliers and affine subspaces with the Frobenius norm reformulation of NSC.

3.1. Dealing with noise

In practice, the data is often contaminated by noise, e.g., Gaussian noise. A common way to deal with Gaussian noise is to relax the equality constraints of (3), yielding the following formulation

$$\min_{\mathbf{C}} \frac{1}{2}\|\mathbf{I} - \mathbf{C}\|_F^2 + \frac{\lambda}{2}\|\mathbf{X}\mathbf{C}\|_F^2 \; . \tag{7}$$

This formulation has the nice property that it has closed-form solution, given simply by solving the linear equation

$$(\mathbf{I} + \lambda\mathbf{X}^T\mathbf{X})\mathbf{C} = \mathbf{I} \; . \tag{8}$$

(a) Motion segmentation (Better viewed in color)

(b) Face clustering

Fig. 1: (a) Motion segmentation: segment the trajectories according to the motions, i.e. the trajectories of the same motion are segmented into the same cluster; (b) Face clustering: find the face images that belong to the same subjects

3.2. Affine Subspace

Affine subspaces are more general subspaces that do not necessarily go through the origin and they are important generalization of linear subspaces.In general, adding affine constraint allows us to represent the subspaces more compactly. So for affine subspaces, we formulate the problem as:

$$\min_{\mathbf{C}} \frac{1}{2}\|\mathbf{I}-\mathbf{C}\|_F^2 + \frac{\lambda}{2}\|\mathbf{X}\mathbf{C}\|_F^2 \quad \text{s.t.} \quad \mathbf{1}^T\mathbf{C} = \mathbf{0} \,. \tag{9}$$

It again has closed-from solution given by solving the linear equation

$$(\nu^T\nu + \lambda\nu^T\mathbf{X}^T\mathbf{X}\nu)\Phi = -\nu^T \,, \tag{10}$$

where ν is the null space of $\mathbf{1}^T$, and Φ is the new variable to estimate. Given Φ, the coefficient matrix $\mathbf{C}$ is computed as $\mathbf{C} = \nu\Phi$.

3.3. Dealing with outliers

When the data is corrupted by sparse outliers, we formulate the problem as

$$\min_{\mathbf{C},\mathbf{E}} \frac{1}{2}\|\mathbf{I}-\mathbf{C}\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{X}\mathbf{C}\|_F^2 + \lambda_2\|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathbf{X}\mathbf{C}-\mathbf{E} = \mathbf{0}, \tag{11}$$

with ℓ_1 norm regularization on the outlier term $\mathbf{E}$.

The Problem (11) can be solved via the Alternating Direction Method of Multipliers (ADMM) [18, 19]. In particular, the corresponding augmented Lagrangian is expressed as:

$$\begin{aligned} L(\mathbf{C},\mathbf{E},\mathbf{Y}) = & \frac{1}{2}\|\mathbf{I}-\mathbf{C}\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{X}\mathbf{C}\|_F^2 + \lambda_2\|\mathbf{E}\|_1 \\ & + \mathrm{tr}\mathbf{Y}^T(\mathbf{X}\mathbf{C}-\mathbf{E}) + \frac{\rho}{2}\|\mathbf{X}\mathbf{C}-\mathbf{E}\|_F^2 \,, \end{aligned} \tag{12}$$

where $\mathbf{Y}$ is the matrix containing the Lagrange multipliers, and ρ is the penalty parameter of the augmented quadratic term.

The detailed algorithm is outlined in Algorithm 1. Specifically, the update of $\mathbf{C}$ can be achieved by solving linear equations, and the update of $\mathbf{E}$ can be obtained by the soft-thresholding operator [19] defined as $\mathcal{T}_\tau[x] = \mathrm{sign}(x) \cdot (\max|x| - \tau, 0)$. Then the update of $\mathbf{E}$ is given by $\mathbf{E} = \mathcal{T}_{\frac{\lambda_2}{\rho}}(\mathbf{X}\mathbf{C} + \mathbf{Y}/\rho)$.

Algorithm 1 Solving (11) via ADMM

Input:
Data matrix $\mathbf{X}$, parameters λ_1, λ_2;

Initialization:
$\mathbf{C} = \mathbf{0}, \mathbf{Y} = \mathbf{0}, \rho = 10^{-6}, \eta > 1, \rho_m = 10^{10}, \varepsilon = 10^{-8}$

while not converged **do**

1. Update $\mathbf{C}$ by solving the linear equation

$$(\mathbf{I} + (\lambda_1+\rho)\mathbf{X}^T\mathbf{X})\mathbf{C} = \mathbf{I} - \mathbf{X}^T\mathbf{Y} + \rho\mathbf{X}^T\mathbf{E} \,.$$

2. Update $\mathbf{E}$ by solving the following problem

$$\begin{aligned} \mathbf{E} &= \operatorname*{argmin}_{\tilde{\mathbf{E}}} \frac{\lambda_2}{\rho}\|\tilde{\mathbf{E}}\|_1 + \frac{1}{2}\|\tilde{\mathbf{E}} - (\mathbf{X}\mathbf{C} + \mathbf{Y}/\rho)\|_F^2 \,, \\ &= \mathcal{T}_{\frac{\lambda_2}{\rho}}(\mathbf{X}\mathbf{C} + \mathbf{Y}/\rho) \,. \end{aligned}$$

3. Update the Lagrange multipliers $\mathbf{Y}$ and the penalty parameter ρ as

$$\begin{aligned} \mathbf{Y} &= \mathbf{Y} + \rho(\mathbf{X}\mathbf{C} - \mathbf{E}) \,, \\ \rho &= \min(\eta\rho, \rho_m) \,. \end{aligned}$$

4. Check convergence

$$\|\mathbf{X}\mathbf{C} - \mathbf{E}\|_\infty < \varepsilon \,.$$

end while

Output: Coefficient matrix $\mathbf{C}$

4. EXPERIMENTS

In this section, we present experiments on two standard datasets, i.e., Hopkins 155 for motion segmentation and Extended Yale B for face clustering, to evaluate the performance of our proposed NSC. We compare NSC with state-of-the-art algorithms, including SSC [8], LRR [9], and LRSC [11]. Note that for the sake of fair comparison, NSC, LRR and LRSC employ the same heuristic way [17] to build the affinity matrix, while SSC builds the affinity matrix as $|\mathbf{C}| + |\mathbf{C}^T|$ since it gets no better result in that heuristic way.

Hopkins 155: Hopkins 155 is a standard motion segmentation dataset with 155 sequences, each of which contains

Table 1: Clustering error (in %) on Hopkins 155.

Method	SSC	LRR	LRSC	NSC
(a) 2F-dimensional data points				
2 motions				
Mean	1.53	2.53	1.69	**1.10**
Median	**0.00**	**0.00**	**0.00**	**0.00**
3 motions				
Mean	4.40	3.69	5.67	**3.07**
Median	0.56	1.10	1.22	**0.20**
All				
Mean	2.18	2.85	2.59	**1.55**
Median	**0.00**	**0.00**	**0.00**	**0.00**
(b) $4k$-dimensional data points by applying PCA				
2 motions				
Mean	1.83	3.67	1.75	**0.85**
Median	**0.00**	**0.00**	**0.00**	**0.00**
3 motions				
Mean	4.40	5.31	5.64	**2.08**
Median	1.53	1.09	1.22	**0.20**
All				
Mean	2.41	4.04	2.63	**1.13**
Median	**0.00**	**0.00**	**0.00**	**0.00**

multiple motions. And each motion corresponds to an affine subspace of dimension three. Since this dataset is free of outliers, we use the formulation (9) for NSC, with $\lambda = 240$. For the baseline algorithms, we've tuned their parameters to the best respectively. The experimental results presented in Table 1 show that NSC outperforms all the other baseline algorithms in terms of clustering accuracy. In terms of runtime, since NSC gets closed-form solution, it's much faster than those iterative algorithms such as SSC and LRR, and is comparable to LRSC which also results in closed form solution.

Extended Yale B: This dataset consists of 38 subjects, each of which contains 64 frontal faces images acquired under different illuminance conditions. We follow the experimental settings of [8] and partition the 38 subjects into four groups (1-10,11-20,21-30, and 31-38). Since this dataset is contaminated with outliers, we use formulation (11) for NSC, with $\lambda_1 = 5 \times 10^{-2}$,$\lambda_2 = 5 \times 10^{-4}$ for $2F$ dimensional data and $\lambda_1 = 6 \times 10^{-3}$,$\lambda_2 = 3 \times 10^{-3}$ for $10k$ dimensional data. The parameters for baseline algorithms have been tuned to the best. Table 2 gives the clustering accuracy for different subjects, and Fig. 2 depicts the runtimes for computing the affinity matrix. Both of them show that NSC outperforms the baselines for the problem of face clustering, which evidences the robustness and efficiency of the proposed algorithm NSC.

5. CONCLUSION

In this paper, we have proposed a novel subspace clustering algorithm called Null Space Clustering and we have employed an equivalent Frobenius norm reformulation which handles the noise and outliers robustly and efficiently. When the data is contaminated with Gaussian noise, our formulation leads to a closed-form solution. When the outliers are present, we efficiently solve the problem via ADMM. We've applied

Table 2: Clustering error (in %) on Extended Yale B.

Method	SSC	LRR	LRSC	NSC
(a)2F-dimensional data points				
2 subjects				
Mean	1.86	2.54	5.32	**1.85**
Median	**0.00**	0.78	4.69	1.56
3 subjects				
Mean	3.10	4.21	8.47	**2.69**
Median	**1.04**	2.60	7.86	2.08
5 subjects				
Mean	4.31	6.90	12.24	**3.42**
Median	**2.50**	5.63	11.25	2.81
8 subjects				
Mean	5.85	14.34	23.72	**4.93**
Median	**4.49**	14.34	28.03	4.59
10 subjects				
Mean	10.94	22.92	30.36	**5.36**
Median	5.63	23.59	28.75	**5.47**
(b)$10k$-dimensional data points by applying PCA				
2 subjects				
Mean	5.91	5.97	3.26	**1.62**
Median	2.34	4.69	2.34	**0.78**
3 subjects				
Mean	9.39	7.77	4.95	**2.13**
Median	6.77	6.25	3.65	**1.56**
5 subjects				
Mean	14.96	12.75	19.46	**2.56**
Median	14.22	12.81	21.56	**2.19**
8 subjects				
Mean	20.43	20.14	39.95	**3.76**
Median	18.36	18.36	40.82	**3.03**
10 subjects				
Mean	23.96	24.74	53.65	**5.63**
Median	24.22	24.38	62.34	**4.53**

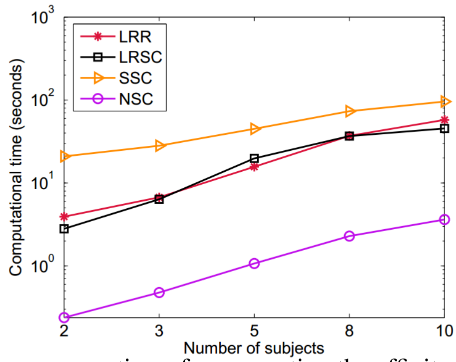

Fig. 2: Average runtimes for computing the affinity matrix on Extended Yale B.

our algorithm for motion segmentation and face clustering and evaluated it on two standard datasets. The experiments have shown that the proposed NSC leads to improved performance in terms of clustering accuracy and efficiency when compared to state-of-the-art algorithms.

Acknowledgements

NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

The research is funded in part by ARC: DP120103896, DP130104567,LP100100588, CE140100016.

6. REFERENCES

[1] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *CVPR*, 2009.

[2] R. Basri and D.W. Jacobs, "Lambertian reflectance and linear subpaces," *IEEE TPAMI*, vol. 25, no. 2, pp. 218–233, 2003.

[3] A. Yang, J. Wright, Y. Ma, and S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *CVIU*, vol. 110, no. 2, pp. 212–225, 2008.

[4] W. Hong, J. Wright, K. Huang, and Y. Ma, "Multi-scale hybrid linear models for lossy image representation," *IEEE TIP*, vol. 15, no. 12, pp. 3655–3671, 2006.

[5] C. W. Gear, "Multibody grouping from motion images," *IJCV*, vol. 29, no. 2, pp. 133–150, 1998.

[6] J.P. Costeira and T. Kanade, "A multibody factorization method for independently moving objects," *IJCV*, vol. 29, no. 3, pp. 159–179, 1998.

[7] R. Vidal, Y. Ma, and S. Sastry, "Generalized principle component analysis(gpca)," *IEEE TPAMI*, vol. 27, no. 12, pp. 1945–1959, 2005.

[8] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE TPAMI*, vol. 35, no. 11, pp. 2765 – 2781, 2013.

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

[10] P. Ji, M. Salzmann, and H. Li, "Efficient dense subspace clustering," in *WACV*, 2014.

[11] R. Vidal and P. Favaro, "Low rank subspace clustering (lrsc)," *Pattern Recognition Letters*, 2013.

[12] S. Rao, A. Yang, S.R Sastry, and Y. Ma, "Robust algebraic segmentation of mixed rigid-body and planar motions in two views," *IJCV*, vol. 88, pp. 425–446, 2010.

[13] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *ICML*, 2010.

[14] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *ECCV*, 2012.

[15] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE TPAMI*, vol. 23, no. 6, pp. 643–660, 2001.

[16] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE TPAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[17] F. Lauer and C. Schnorr, "Spectral clustering of linear subspaces for motion segmentation," in *CVPR*, 2009.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[19] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *UIUC Technical Report*, vol. UILU-ENG-09-2215, 2009.