On Sufficient Condition for Affine Sparse Subspace Clustering

Chun-Guang Li
SICE, Beijing University of Posts and Telecommunications
Beijing 100876, P.R. China

Chong You and René Vidal
Center for Imaging Science, Johns Hopkins University
Baltimore, MD 21218, USA

In this paper, we consider the following problem:

Problem 1 (affine subspace clustering). *Let $X \in I\!R^{D\times N}$ be a real-valued matrix whose columns are drawn from a union of n affine subspaces of $I\!R^D$, $\bigcup_{\ell=1}^n\{\mathcal{A}_\ell\}$, of dimension $d_\ell < D$, for $\ell = 1,\ldots,n$. The goal of affine subspace clustering is to segment the columns of X into their corresponding affine subspaces.*

One way to approach this problem is by the Sparse Subspace Clustering (SSC) algorithm [1], which is based on the idea that each point $\boldsymbol{x}_j$ in a linear subspace of dimension d_ℓ can be written as a linear combination of (at most d_ℓ) other points in its own subspace. Although primarily designed for linear subspaces, SSC has been applied to affine subspaces by further requiring that the coefficients add up to 1. This leads to the following optimization problem

$$\min_{c_j} \|\boldsymbol{c}_j\|_1, \quad \text{s.t.} \quad \boldsymbol{x}_j = X\boldsymbol{c}_j, \quad c_{jj} = 0, \quad \boldsymbol{1}^\top \boldsymbol{c}_j = 1, \tag{1}$$

where $\boldsymbol{c}_j \in I\!R^N$ is the vector of coefficients and $\boldsymbol{1}$ is the vector of all ones with appropriate dimension. Then spectral clustering is applied to the induced affinity. We call this subspace clustering approach with modified formulation (1) as *Affine Sparse Subspace Clustering* (ASSC). One expects that the sparsest affine representation $\boldsymbol{c}_j$ selects only data points from the affine subspace which $\boldsymbol{x}_j$ belongs to. This property can be captured by the definition below.

Definition 1 (ℓ_1 affine subspace detection property). *A data point $\boldsymbol{x}_j \in \mathcal{X}$ in an affine subspace $\mathcal{A}_\ell$ obeys the ℓ_1 Affine Subspace Detection Property (ℓ_1-ASDP) if and only if it holds that the optimal solution to problem* (1) *has nonzero entries corresponding only to data points from the correct affine subspace $\mathcal{A}_\ell$.*

Prior work on theoretical analysis of SSC considers mostly the case of *linear* subspaces and has established sufficient geometric conditions for the optimal solution of the standard problem to be such that its nonzero entries correspond to points in the correct subspace [1], [3]. However, there is a lack of theoretical understanding of the behavior of ASSC for affine subspaces. In this paper, we present a geometric analysis of ASSC and give a deterministic sufficient condition under which it succeeds to meet ℓ_1-ASDP. Moreover we reveal the *curse* and the *blessing* from the affine constraint – that might deepen our understanding of the behaviors of ASSC.

To facilitate the analysis, we reformulate ASSC as follows

$$\min_{\boldsymbol{c}_j} \|\boldsymbol{c}_j\|_1, \quad \text{s.t.} \quad Y_{-j}\boldsymbol{c}_j = 0, \quad \boldsymbol{1}^\top \boldsymbol{c}_j = 1, \tag{2}$$

in which $\boldsymbol{c}_j \in I\!R^{N-1}$ and $Y_{-j} = X_{-j} - \boldsymbol{x}_j\boldsymbol{1}^\top$ where X_{-j} is the matrix X with the j-th column removed. Consider a data point $\boldsymbol{x}_j$ lying in affine subspace $\mathcal{A}_\ell$. We arrange data points $\{\boldsymbol{x}_i |\ i \neq j, \boldsymbol{x}_i \in \mathcal{A}_\ell\}$ as matrix $X_{-j}^{(\ell)}$, data points $\{\boldsymbol{x}_i - \boldsymbol{x}_j |\ i \neq j, \boldsymbol{x}_i \in \mathcal{A}_\ell\}$ as matrix $Y_{-j}^{(\ell)}$, and data points $\{\boldsymbol{x}_i - \boldsymbol{x}_j |\ \boldsymbol{x}_i \in \mathcal{A}_\kappa,\ \kappa \neq \ell\}$ as matrix $Y_{-j}^{(\kappa)}$. Note that by subtracting $\boldsymbol{x}_j$, the affine subspace $\mathcal{A}_\ell$ reduces to *linear* subspace $\mathcal{S}_\ell$, i.e., the columns of $Y_{-j}^{(\ell)}$ lie in *linear* subspace $\mathcal{S}_\ell$. Let $U^{(\ell)} \in I\!R^{D\times d_\ell}$ be an orthogonal basis for linear subspace $\mathcal{S}_\ell$, then we have that $A_{-j}^{(\ell)} = U^{(\ell)\top}Y_{-j}^{(\ell)}$. To define the affine subspace incoherence, we consider an optimization problem as follows

$$\min_{\boldsymbol{c}} \|\boldsymbol{c}\|_1 \quad \text{s.t.} \quad A_{-j}^{(\ell)}\boldsymbol{c} = \boldsymbol{0}, \quad \boldsymbol{1}^\top\boldsymbol{c} = 1. \tag{3}$$

The Lagrangian dual is as follows

$$\max_{w,\nu} \quad -\nu \quad \text{s.t.} \quad \|A_{-j}^{(\ell)\top}\boldsymbol{w} + \nu\boldsymbol{1}\|_\infty \leq 1. \tag{4}$$

where $\boldsymbol{w} \in I\!R^D$ and $\nu \in I\!R$ are dual variables. Denote the optimal solution of problem (4) as $(\boldsymbol{w}_j^*, \nu_j^*)$. We define $(\boldsymbol{v}_j^*, \nu_j^*)$ as a *dual point* of data point $\boldsymbol{x}_j$ where $\boldsymbol{v}_j^* = U^{(\ell)}\boldsymbol{w}_j^*$, and $\boldsymbol{q}_j = \frac{\tilde{\boldsymbol{v}}_j^*}{\|\tilde{\boldsymbol{v}}_j^*\|_2}$ as the *augmented dual direction* with respect to the dual point $(\boldsymbol{v}_j^*, \nu_j^*)$ where $\boldsymbol{q}_j \in I\!R^{D+1}$ and $\tilde{\boldsymbol{v}}_j^* = [\boldsymbol{v}_j^{*\top}, \nu_j^*]^\top$.

Definition 2 (affine subspace incoherence). *The affine subspace incoherence of a point $\boldsymbol{x}_j \in \mathcal{A}_\ell$ vis a vis the other points $\boldsymbol{x} \in \mathcal{A}_\kappa(\kappa \neq \ell)$ is defined as follows $\tilde{\mu}_j \doteq \max\{\|\tilde{Y}_{-j}^{(\kappa)\top}\boldsymbol{q}_j\|_\infty, \quad \kappa = 1,\cdots,n, \quad \kappa \neq \ell\}$ where $\tilde{Y}_{-j}^{(\kappa)} = [Y_{-j}^{(\kappa)\top}, \boldsymbol{1}]^\top$.*

Theorem 1 (ℓ_1 affine subspace separation theorem). *Given a data set $\mathcal{X}$, if the following affine subspace incoherence condition*

$$\tilde{\mu}_j < r(\tilde{\mathcal{P}}_j^\ell), \tag{5}$$

is satisfied for all data points $\boldsymbol{x}_j \in \mathcal{X}$, where $\tilde{\mathcal{P}}_j^\ell$ is the symmetric convex hull of columns of $\tilde{Y}_{-j}^{(\ell)}$ and $r(\tilde{\mathcal{P}}_j^\ell)$ is the inradius of $\tilde{\mathcal{P}}_j^\ell$. Then, ℓ_1-ASDP holds.

Note that the affine constraint in ASSC may abrogate the sparsity-promoting property of ℓ_1-norm when the feasible solution $\boldsymbol{c}$ is nonnegative, because $\|\boldsymbol{c}\|_1 = \boldsymbol{1}^\top\boldsymbol{c} = 1$ and thus it is of no effect to penalize the solution. Our analysis show that the affine constraint may bring a *curse* and a *blessing* to ASSC. The curse is that ASSC fails to meet the ℓ_1-ASDP if the convex hull surrounding $\boldsymbol{x}_j$ can be formed by data points not only from the correct affine subspace. The blessing is that the connectivity issue [2] of SSC might be partially alleviated.

Theorem 2 (failure in intersection). *ASSC fails to meet ℓ_1-ASDP for $\boldsymbol{x}_j \in \mathcal{A}_\ell$ if conv$\{Y_{-j}^{(\ell)}\}$ and conv$\{Y_{-j} \setminus Y_{-j}^{(\ell)}\}$ intersect.*

Theorem 3 (grouping effect for interior points). *For a point $\boldsymbol{x}_i$ in $\mathcal{A}_\ell$, suppose that the condition* (5) *is satisfied so that ℓ_1-ASDP holds. If $\boldsymbol{x}_i$ is in the relative interior of conv$\{X_{-i}^{(\ell)}\}$, then there exist an optimal solution to* (1) *which is subspace-dense, i.e., the coefficients for all other points in $\mathcal{A}_\ell$ are nonzero.*

Acknowledgments

C.-G. Li was partially supported by National Natural Science Foundation of China under Grant Nos. 61273217 and 61175011, and Scientific Research Foundation for the Returned Overseas Chinese Scholars, Ministry of Education of China. C. You and R. Vidal were supported by the National Science Foundation under Grant No. 1447822.

REFERENCES

[1] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

[2] B. Nasihatkon and R. Hartley. Graph connectivity in sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[3] M. Soltanolkotabi and E. J. Candès. A geometric analysis of subspace clustering with outliers. *Annals of Statistics*, 40(4):2195–2238, 2012.