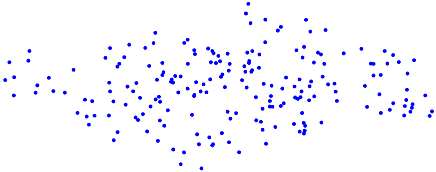
A Geometric Analysis of Subspace Clustering with Outliers

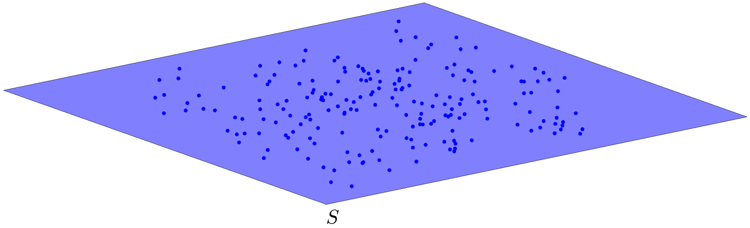
Mahdi Soltanolkotabi and Emmanuel Candés

Stanford University

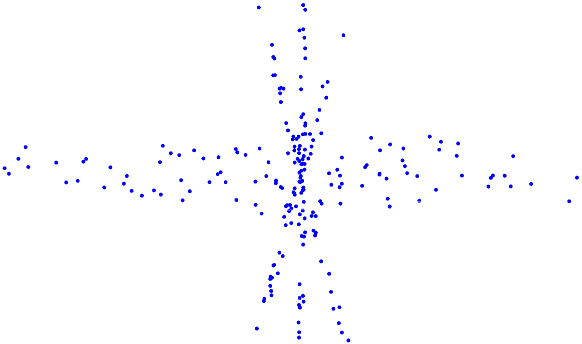
Fundamental Tool in Data Mining : PCA



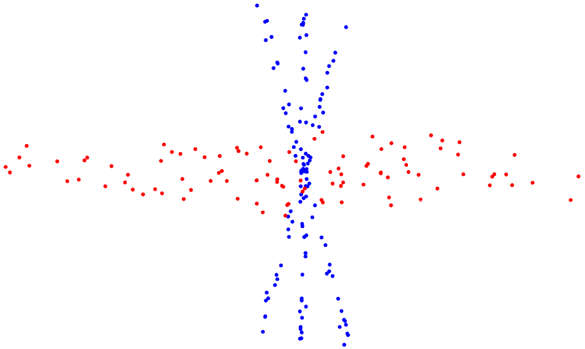
Fundamental Tool in Data Mining : PCA



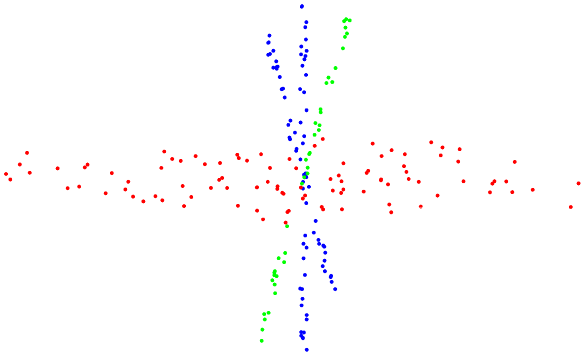
Subspace Clustering : multi-subspace PCA



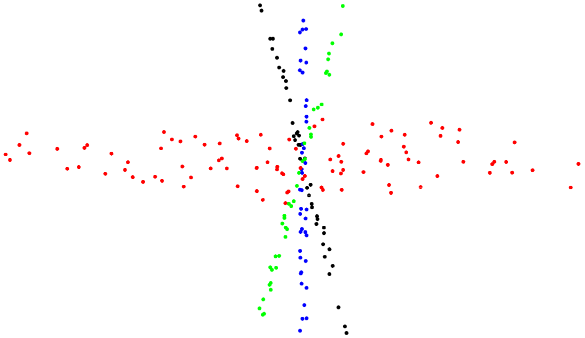
Subspace Clustering : multi-subspace PCA



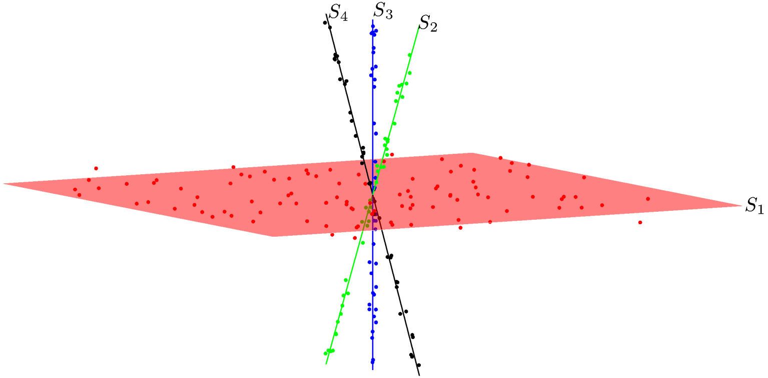
Subspace Clustering : multi-subspace PCA



Subspace Clustering : multi-subspace PCA



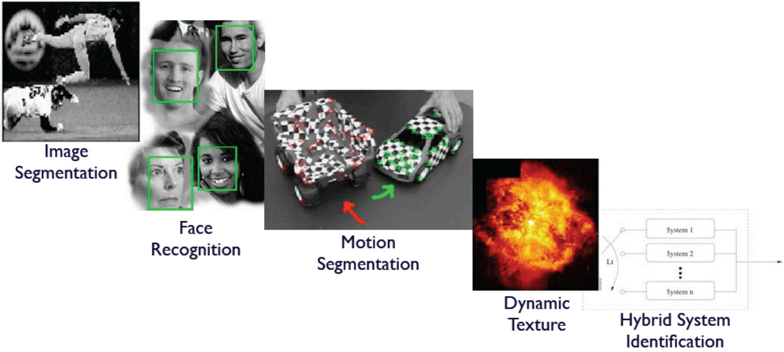
Subspace Clustering : multi-subspace PCA



Index

Computer Vision

- Many applications.

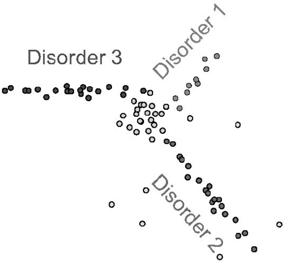


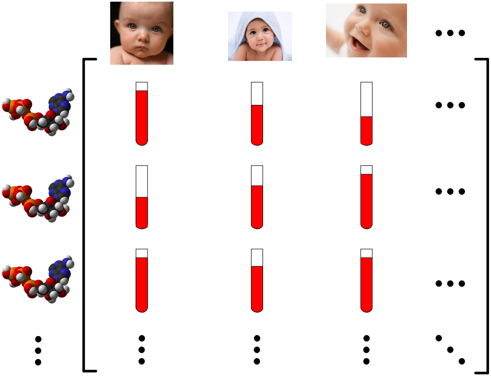
Figures from various papers/talks.

- I refer to the interesting lecture by Prof. Ma tomorrow which highlights the relevance of low dimensional structures in computer vision. Just add multiple categories :)

Metabolic Screening

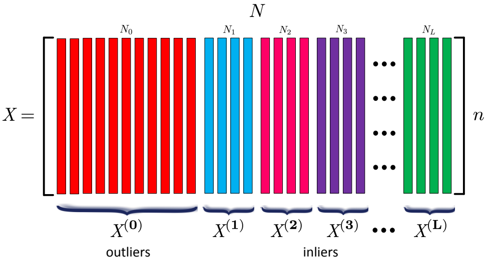
- **Goal** Detect metabolic disease as early as possible in new borns.
- concentration of metabolites are tested.
- Usually, each metabolic disease causes a correlation between the concentration of a specific set of metabolites.





Problem Formulation

- N unit norm points in $\mathbb{R}^n$ on a union of unknown linear subspaces $S_1 \cup S_2 \cup \ldots \cup S_L$ of unknown dimensions $d_1, d_2, \ldots, d_L$. Also, N_0 uniform outliers.

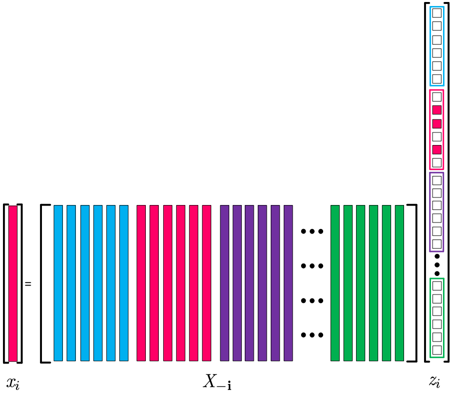


- **Goal** : Without any prior knowledge about the number of subspaces, their orientation or their dimension,
  1) identify all the outliers, and
  2) segment or assign each data point to a cluster as to recover all the hidden subspaces.

Sparse Subspace Clustering

- Regress one column against other columns.
- for $i = 1, \ldots, N$ solve

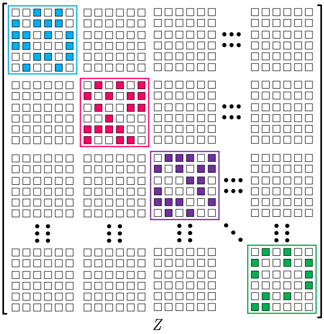
$$\min \|\boldsymbol{z}_i\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{z}_i, \quad \boldsymbol{z}_{ii} = 0.$$



ℓ_1 subspace detection property

Definition (ℓ_1 Subspace Detection Property)

The subspaces $\{S_\ell\}_{\ell=1}^{L}$ and points $\boldsymbol{X}$ obey the ℓ_1 subspace detection property if and only if it holds that for all i, the optimal solution has nonzero entries only when the corresponding columns of $\boldsymbol{X}$ are in the same subspace as $\boldsymbol{x_i}$.



SSC Algorithm

Algorithm 1 Sparse Subspace Clustering (SSC)

Input: A data set $\mathcal{X}$ arranged as columns of $\boldsymbol{X} \in \mathbb{R}^{n \times N}$.

1. Solve (the optimization variable is the $N \times N$ matrix $\boldsymbol{Z}$)

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{Z}\|_{\ell_1} \\ \text{subject to} & \boldsymbol{X}\boldsymbol{Z} = \boldsymbol{X} \\ & \operatorname{diag}(\boldsymbol{Z}) = \mathbf{0}. \end{array}$$

2. Form the affinity graph G with nodes representing the N data points and edge weights given by $\boldsymbol{W} = |\boldsymbol{Z}| + |\boldsymbol{Z}|^T$.
3. Sort the eigenvalues $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$ of the normalized Laplacian of G in descending order, and set

$$\hat{L} = N - \underset{i=1,\ldots,N-1}{\arg\max}\ \sigma_i - \sigma_{i+1}.$$

4. Apply a spectral clustering technique to the affinity graph using $\hat{L}$.

Output: Partition $\mathcal{X}_1, \ldots, \mathcal{X}_{\hat{L}}$.

History of subspace problems and use of ℓ_1 regression

- Various subspace clustering algorithms proposed by Profs. Ma, Sastry, Vidal, Wright and collaborators.

History of subspace problems and use of ℓ_1 regression

- Various subspace clustering algorithms proposed by Profs. Ma, Sastry, Vidal, Wright and collaborators.
- Sparse representation in the multiple subspace framework initiated by Prof. Yi Ma and collaborators in the context of face recognition.

History of subspace problems and use of ℓ_1 regression

- Various subspace clustering algorithms proposed by Profs. Ma, Sastry, Vidal, Wright and collaborators.
- Sparse representation in the multiple subspace framework initiated by Prof. Yi Ma and collaborators in the context of face recognition.
    - Different from subspace clustering because you know the subspaces in advance.

History of subspace problems and use of ℓ_1 regression

- Various subspace clustering algorithms proposed by Profs. Ma, Sastry, Vidal, Wright and collaborators.
- Sparse representation in the multiple subspace framework initiated by Prof. Yi Ma and collaborators in the context of face recognition.
    - Different from subspace clustering because you know the subspaces in advance.
- For subspace clustering by Elhamifar and Vidal. Coined the name Sparse Subspace Clustering (SSC).

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contribution is to bridge the gap between theory and practice, specifically :

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contribution is to bridge the gap between theory and practice, specifically :

- Intersecting Subspaces.

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contribution is to bridge the gap between theory and practice, specifically :

- Intersecting Subspaces.

- Subspaces of nearly linear dimension.

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contribution is to bridge the gap between theory and practice, specifically :

- Intersecting Subspaces.

- Subspaces of nearly linear dimension.

- Outliers.

Our goal

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice. Limited theoretical guarantees, Huge gap between theory and practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contribution is to bridge the gap between theory and practice, specifically :

- Intersecting Subspaces.

- Subspaces of nearly linear dimension.

- Outliers.

Models

We will consider three different models regarding the orientation of subspaces and distribution of points on each subspace.

	Orientation	distribution
Deterministic		
Semi-random		
Fully-random		

: Random : Deterministic

Deterministic Model

Theorem (M. Soltanolkotabi, E. J. Candes)

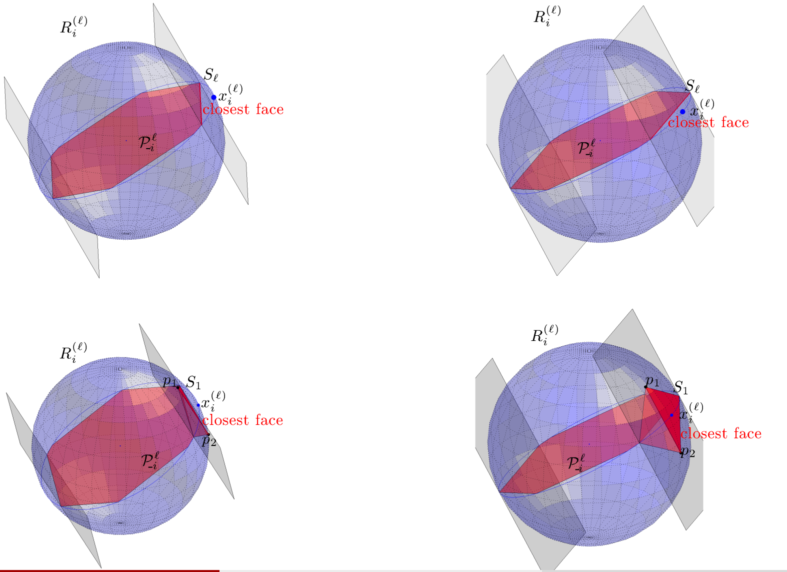
If

$$\mu(\mathcal{X}_\ell) < \min_{i:\ \boldsymbol{x}_i \in \mathcal{X}_\ell} r(\mathcal{P}^\ell_{-i}) \tag{1}$$

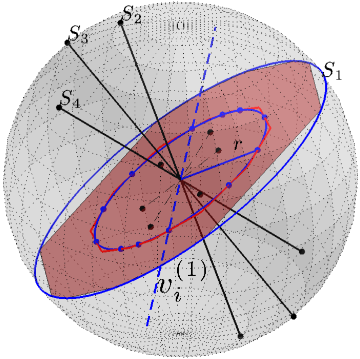
for each $\ell = 1, \ldots, L$, then the subspace detection property holds. If (1) holds for a given ℓ, then a local subspace detection property holds in the sense that for all $\boldsymbol{x}_i \in \mathcal{X}_\ell$, the solution to the optimization problem has nonzero entries only when the corresponding columns of $\boldsymbol{X}$ are in the same subspace as $\boldsymbol{x_i}$.

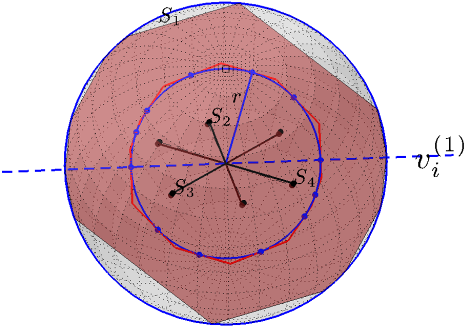
As long as the point sets are not very "similar" and the points on each subspace are well distributed SSC works.

Geometric view of ℓ_1 subspace detection



Geometric view on the condition of the theorem





principal angle and affinity between subspaces

Definition

The principal angles $\theta_{k,\ell}^{(1)}, \ldots, \theta_{k,\ell}^{(d_k \bigvee d_\ell\}}$ between two subspaces S_k and S_ℓ of dimensions d_k and d_ℓ, are recursively defined by

$$\cos(\theta_{k\ell}^{(i)}) = \max_{\boldsymbol{y}\in S_k} \max_{\boldsymbol{z}\in S_\ell} \frac{\boldsymbol{y}^T\boldsymbol{z}}{\|\boldsymbol{y}\|_{\ell_2}\|\boldsymbol{z}\|_{\ell_2}} = \frac{\boldsymbol{y}_i^T\boldsymbol{z}_i}{\|\boldsymbol{y}_i\|_{\ell_2}\|\boldsymbol{z}_i\|_{\ell_2}}.$$

with the orthogonality constraints $\boldsymbol{y}^T\boldsymbol{y}_j = 0$, $\boldsymbol{z}^T\boldsymbol{z}_j = 0$, $j = 1, \ldots, i-1$.

Definition

The affinity between two subspaces is defined by

$$\textit{aff}(S_k, S_\ell) = \sqrt{\cos^2\theta_{k\ell}^{(1)} + \ldots + \cos^2\theta_{k\ell}^{(d_k \bigvee d_\ell)}}.$$

For random subspaces square of affinity is the Pillai-Bartlett test Statistics.

Semi-random model

Assume we have $N_\ell = \rho_\ell d_\ell + 1$ random points on subspace S_ℓ.

Theorem (M. Soltanolkotabi, E. J. Candes)

Ignoring some log factors, as long as

$$\max_{k\,:\,k\neq\ell} \frac{\mathrm{aff}(S_k, S_\ell)}{\sqrt{d_k}} < c\sqrt{\log \rho_\ell}, \quad \textit{for each } \ell, \tag{2}$$

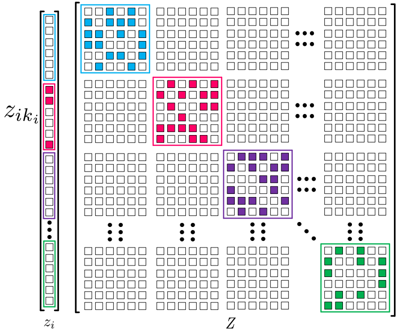
the subspace detection property holds with high probability.

- The affinity can at most be $\sqrt{d_k}$ and, therefore, our result essentially states that if the affinity is less than $c\sqrt{d_k}$, then SSC works.
- allows for intersection of subspaces.

Four error criterion

- **feature detection error** : For each point $\boldsymbol{x}_i$ in subspace S_{ℓ_i} we partition the optimal solution of SSC ($\boldsymbol{z}_i$) in the form

$$\boldsymbol{z}_i = \boldsymbol{\Gamma} \begin{bmatrix} \boldsymbol{z}_{i1} \\ \boldsymbol{z}_{i2} \\ \boldsymbol{z}_{i3} \end{bmatrix}.$$



$$\frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{\|\boldsymbol{z}_{ik_i}\|_{\ell_1}}{\|\boldsymbol{z}_i\|_{\ell_1}}\right).$$

- **clustering error** :

$$\frac{\#\text{ of misclassified points}}{\text{total }\#\text{ of points}}.$$

Four error criterion

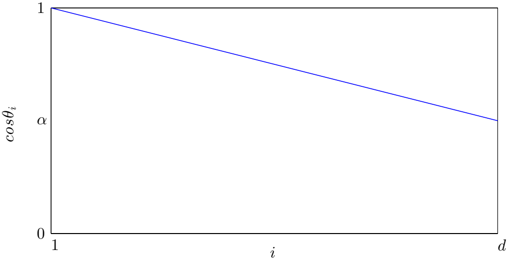
- **error in estimating the number of subspaces :** This error quantity indicates whether our version of the SSC algorithm can correctly identify the correct number of subspaces ; 0 when it succeeds, 1 when it fails.
- **Singular values of the normalized Laplacian :**

$$\sigma_{N-L} < \epsilon.$$

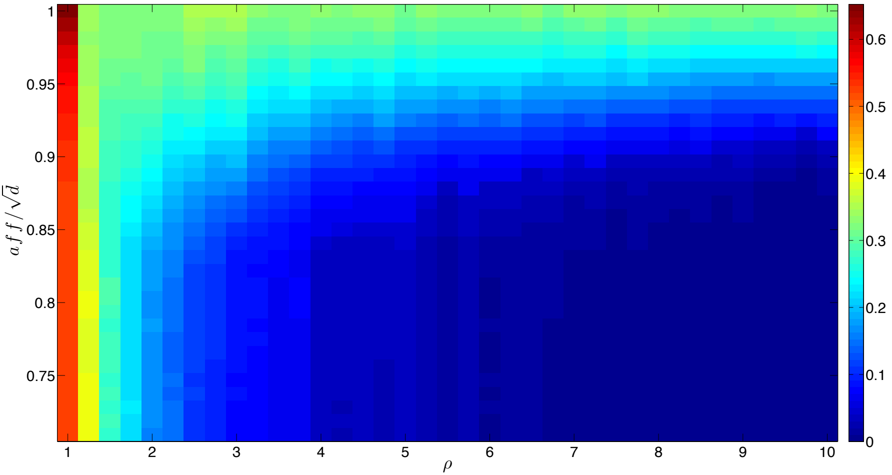
affinity example

Choose difficult case of $n = 2d$, the three basis we choose for each subspace are

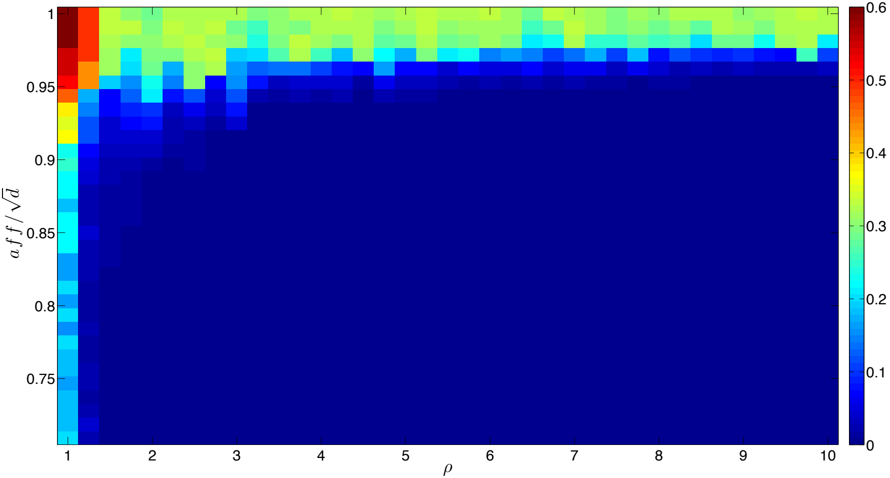
$$\boldsymbol{U}^{(1)} = \begin{bmatrix} \boldsymbol{I}_d \\ \boldsymbol{0}_{d\times d} \end{bmatrix}, \quad \boldsymbol{U}^{(2)} = \begin{bmatrix} \boldsymbol{0}_{d\times d} \\ \boldsymbol{I}_d \end{bmatrix}, \boldsymbol{U}^{(3)} = \begin{bmatrix} \cos(\theta_1) & 0 & 0 & 0 & \dots & 0 \\ 0 & \cos(\theta_2) & 0 & 0 & \dots & 0 \\ 0 & 0 & \cos(\theta_3) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \cos(\theta_d) \\ \sin(\theta_1) & 0 & 0 & 0 & \dots & 0 \\ 0 & \sin(\theta_2) & 0 & 0 & \dots & 0 \\ 0 & 0 & \sin(\theta_3) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \sin(\theta_d) \end{bmatrix}.$$



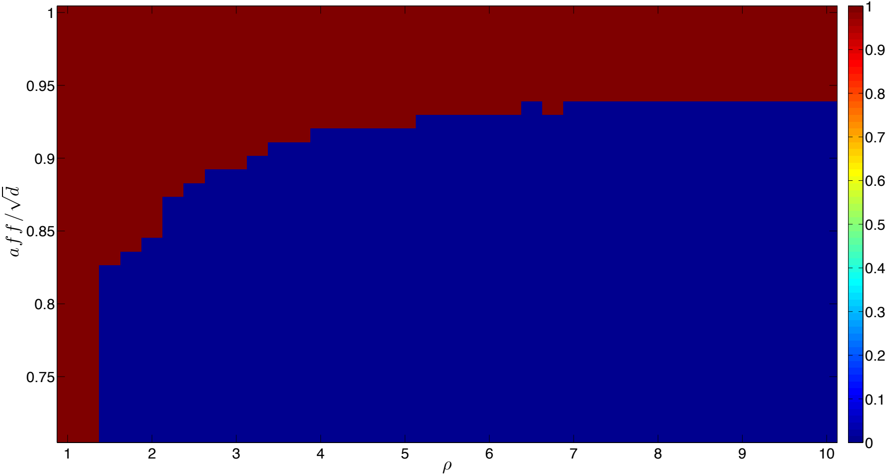
feature detection error



clustering error



error in estimating the number of subspaces



Fully-random model

L subspaces, each of dimension d, $\rho d + 1$ points on each subspace.

Theorem (M. Soltanolkotabi, E. J. Candés)

The subspace detection property holds with high probability as long as

$$d < c\, n,$$

where $c > \frac{c^2(\rho)\log\rho}{12\log N}$.

Fully-random model

L subspaces, each of dimension d, $\rho d + 1$ points on each subspace.

Theorem (M. Soltanolkotabi, E. J. Candés)

The subspace detection property holds with high probability as long as

$$d < c\, n,$$

where $c > \frac{c^2(\rho)\log\rho}{12\log N}$.

- allows for the dimension of the subspaces d to grow almost linearly with the ambient dimension n.

Fully-random model

L subspaces, each of dimension d, $\rho d + 1$ points on each subspace.

Theorem (M. Soltanolkotabi, E. J. Candés)

The subspace detection property holds with high probability as long as

$$d < c\, n,$$

where $c > \frac{c^2(\rho)\log \rho}{12 \log N}$.

- allows for the dimension of the subspaces d to grow almost linearly with the ambient dimension n.
- tend $n \to \infty$, with $\frac{d}{n}$ a constant, $\rho = d^\eta$ with $\eta > 0$ the condition reduces to

$$d < \frac{1}{48(\eta + 1)} n$$

Comparison with previous analysis

Condition of Elhamifar and Vidal

$$\max_{k:\, k\neq\ell} \cos(\theta_{k\ell}) < \frac{1}{\sqrt{d_\ell}} \max_{\boldsymbol{Y}\in\mathbb{W}_{d_\ell}(\boldsymbol{X}^{(\boldsymbol{\ell})})} \sigma_{\min}(\boldsymbol{Y}) \quad \text{for all } \ell = 1,\ldots,L,$$

- **semi-random model :**
  -Elhamifar and Vidal's condition reduces to $\cos\theta_{k\ell} < \frac{1}{\sqrt{d}}$.
  -We require

  $$\text{aff}(S_k, S_\ell) = \sqrt{\cos^2(\theta_{k\ell}^{(1)}) + \cos^2(\theta_{k\ell}^{(2)}) + \ldots + \cos^2(\theta_{k\ell}^{(d)})} < c\sqrt{d}.$$

  The left-hand side can be much smaller than $\sqrt{d}\cos\theta_{k\ell}$ and is, therefore, less restrictive, e.g. $\cos\theta_{k\ell} < c$ is sufficient.

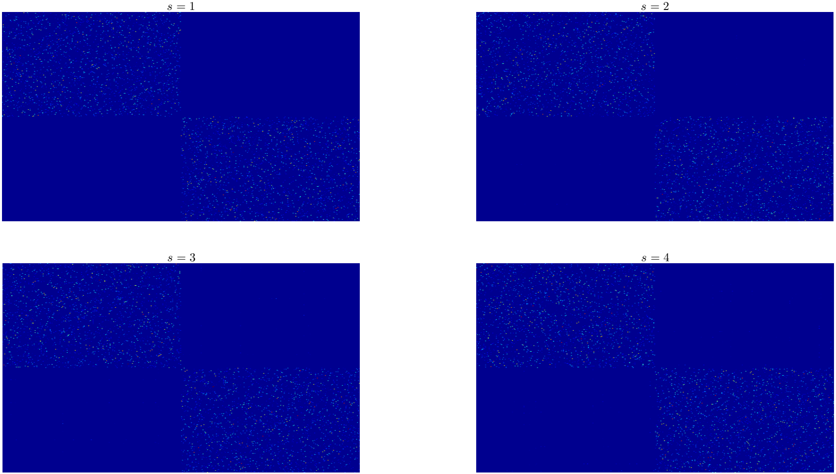
Comparison with previous analysis contd.

- **Intersecting subspaces :**
  Two subspaces with an intersection of dimension s.
  -Elhamifar and Vidal becomes $1 < \frac{1}{\sqrt{d}}$, which cannot hold.
  -Our condition simplifies to

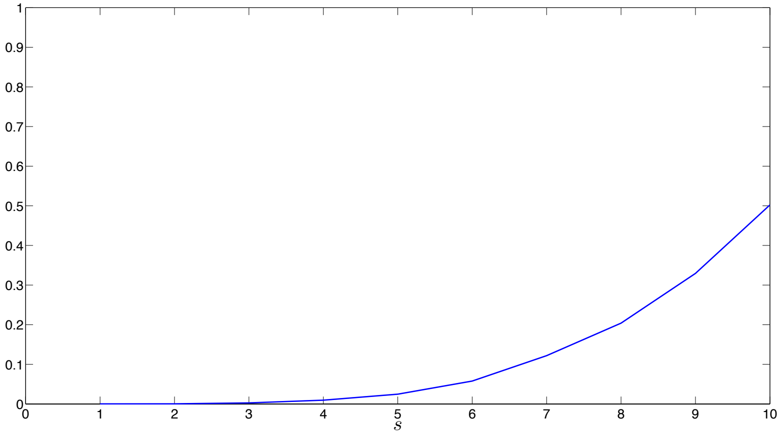
$$\cos^2(\theta_{k\ell}^{(s+1)}) + \ldots + \cos^2(\theta_{k\ell}^{(d)}) < cd - s,$$

- **fully random model :**
  -Elhamifar and Vidal : $d < c_1\sqrt{n}$.
  - We essentially require : $d < cn$.

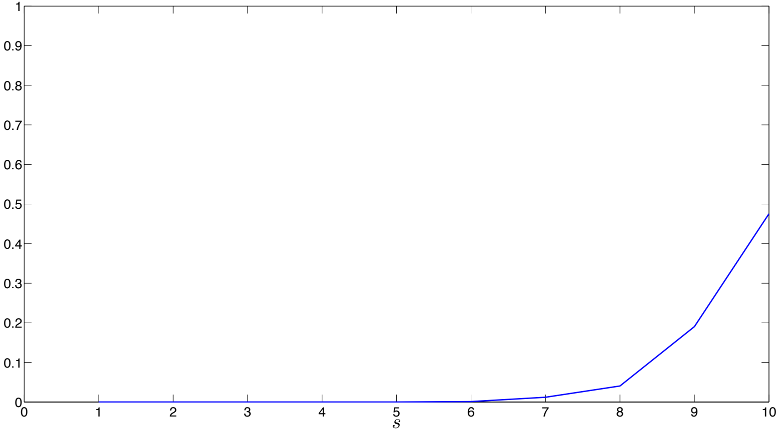
Subspaces Intersect



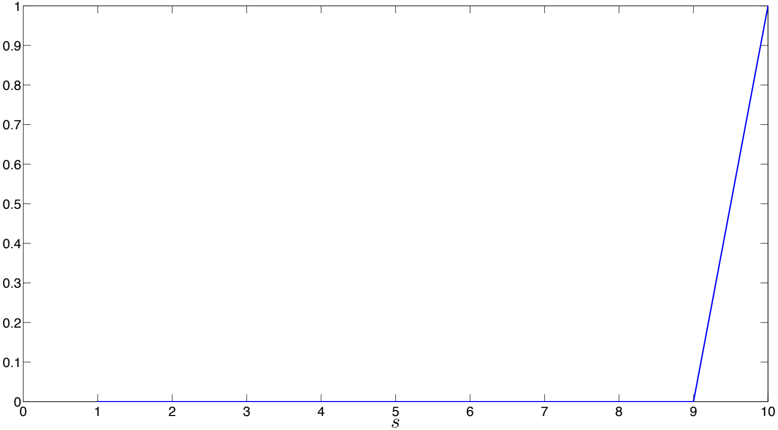
feature detection error



clustering error



error in estimating the number of subspaces



subspace clustering with outliers

Assume the outliers are uniform at random on the unit sphere. Again, regress one column against other columns.
for $i = 1, \ldots, N$ solve

$$\min \|\boldsymbol{z}_i\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{z}_i, \quad \boldsymbol{z}_{ii} = 0.$$

consider optimal values $\|\boldsymbol{z}_i^*\|_{\ell_1}$.

- Inliers : Intuitively, size of support d, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{d}$.

subspace clustering with outliers

Assume the outliers are uniform at random on the unit sphere. Again, regress one column against other columns.
for $i = 1, \ldots, N$ solve

$$\min \|\boldsymbol{z}_i\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{z}_i, \quad \boldsymbol{z}_{ii} = 0.$$

consider optimal values $\|\boldsymbol{z}_i^*\|_{\ell_1}$.

- Inliers : Intuitively, size of support d, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{d}$.
- Outliers : Intuitively, size of support n, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{n}$.

subspace clustering with outliers

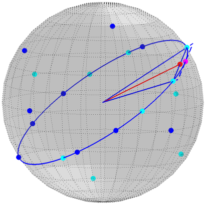
Assume the outliers are uniform at random on the unit sphere. Again, regress one column against other columns.
for $i = 1, \ldots, N$ solve

$$\min \|\boldsymbol{z}_i\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{z}_i, \quad \boldsymbol{z}_{ii} = 0.$$

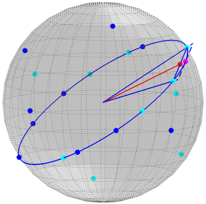
consider optimal values $\|\boldsymbol{z}_i^*\|_{\ell_1}$.

- Inliers : Intuitively, size of support d, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{d}$.
- Outliers : Intuitively, size of support n, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{n}$.
- Therefore, we should expect a gap.

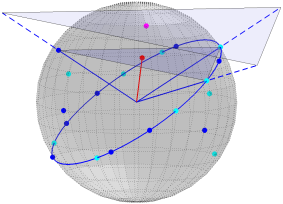
- Inliers : Intuitively, size of support d, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{d}$. Here, $\|\boldsymbol{z}_i^*\|_{\ell_1} = 1.05$.



- Inliers : Intuitively, size of support d, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{d}$. Here, $\|\boldsymbol{z}_i^*\|_{\ell_1} = 1.05$.



- Outliers : Intuitively, size of support n, $\|\boldsymbol{z}_i^*\|_{\ell_1} \sim \sqrt{n}$. Here, $\|\boldsymbol{z}_i^*\|_{\ell_1} = 1.72$.



Algorithm for subspace clustering with outliers [M. Soltanolktoabi, E. J. Candes]

Algorithm 2 Subspace clustering in the presence of outliers

Input: A data set $\mathcal{X}$ arranged as columns of $\boldsymbol{X} \in \mathbb{R}^{n \times N}$.

1. Solve

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{Z}\|_{\ell_1} \\ \text{subject to} & \boldsymbol{X}\boldsymbol{Z} = \boldsymbol{X} \\ & \operatorname{diag}(\boldsymbol{Z}) = \mathbf{0}. \end{array}$$

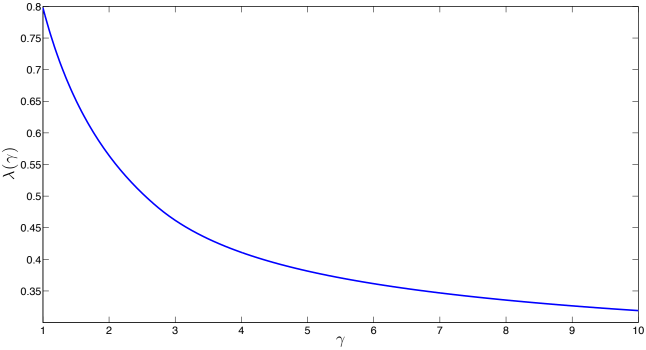
2. For each $i \in \{1, \ldots, N\}$, declare i to be an outlier iff $\|\boldsymbol{z}_i\|_{\ell_1} > \lambda(\gamma)\sqrt{n}$.[3]
3. Apply a subspace clustering to the remaining points.

Output: Partition $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_L$.

2 $\gamma = \frac{N-1}{n}$ and λ is a threshold ratio function.

Threshold function

$$\lambda(\gamma) = \begin{cases} \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\gamma}}, & \text{if } 1 \leq \gamma \leq e, \\ \sqrt{\frac{2}{\pi e}} \frac{1}{\sqrt{\log \gamma}}, & \text{if } \quad \gamma \geq e, \end{cases}$$



with outliers

Theorem (M. Soltanolkotabi, E. J. Candés)

With outlier points uniformly random using the threshold value $(1-t)\frac{\lambda(\gamma)}{\sqrt{e}}\sqrt{n}$, all outliers are identified correctly with high probability. Furthermore, we have the following guarantees in the deterministic and semi-random models.

(a) *If in the deterministic model,*

$$\max_{\ell,i} \frac{1}{r(\mathcal{P}(\boldsymbol{X}_{-i}^{(\ell)}))} < (1-t)\frac{\lambda(\gamma)}{\sqrt{e}}\sqrt{n}, \tag{3}$$

then no 'real' data point is wrongfully detected as an outlier.

(b) *If in the semi-random model,*

$$\max_{\ell} \frac{\sqrt{2d_\ell}}{c(\rho_\ell)\sqrt{\log \rho_\ell}} < (1-t)\frac{\lambda(\gamma)}{\sqrt{e}}\sqrt{n}, \tag{4}$$

then w.h.p. no 'real' data point is wrongfully detected as an outlier.

Comparison with previous analysis

our results restricts the number of outliers to :

$$N_0 < \min\{\rho^{c_2 n/d}, \frac{1}{n} e^{c\sqrt{n}}\} - N_d$$

Lerman and Zhang use

$$e_{\ell_p}(\mathcal{X}, S_1, \ldots, S_L) = \sum_{\boldsymbol{x} \in \mathcal{X}} \min_{1 \le \ell \le L} \big(\mathsf{dist}(\boldsymbol{x}, S_\ell)\big)^p.$$

with $0 \le p \le 1$. They bound on the number of outliers in the semi-random model :

$$N_0 < \tau_0 \rho d \min\Big(1, \min_{k \neq \ell} \mathsf{dist}(S_k, S_\ell)^p / 2^p\Big).$$

which is upper bounded by ρd, the typical number of points per subspace.
non-convex, no practical method known for solving it.

Segmentation with outliers

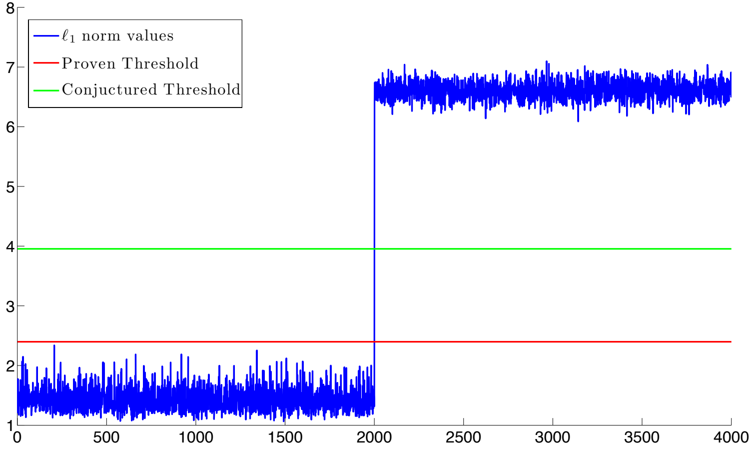


Figure: $d = 5$, $n = 200$, $\rho = 5$, $L = 80$

References

List is not comprehensive!

- A geometric analysis of subspace clustering. M. Soltanolkotabi and E. J. Candes to appear in Annals of Statistics.
- Noisy case, in preparation. Joint with E. J. Candes, E. Elhamifar, and R. Vidal.
- `www.stanford.edu/~mahdisol`.
- Sparse Subspace Clustering. E. Elhamifar and R. Vidal.
- Subspace Clustering. Turorial by R. Vidal.
- Prof. Yi Ma's website for papers on related topics : `http://yima.csl.illinois.edu/`.

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.
- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contributions is to bridge the gap between theory and practice, specifically :

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.
- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contributions is to bridge the gap between theory and practice, specifically :
- Intersecting Subspaces.

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contributions is to bridge the gap between theory and practice, specifically :

- Intersecting Subspaces.

- Subspaces of nearly linear dimension.

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.
- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contributions is to bridge the gap between theory and practice, specifically :
- Intersecting Subspaces.
- Subspaces of nearly linear dimension.
- Outliers.

Conclusion

- Prior work restrictive, essentially say that the subspaces must be almost independent. Can not fully explain why SSC works so well in practice.

- In direct analogy with ℓ_1 regression results before compressed sensing results in 2004.

  Our contributions is to bridge the gap between theory and practice, specifically :

- Intersecting Subspaces.

- Subspaces of nearly linear dimension.

- Outliers.

- Geometric Insights.

Thank you