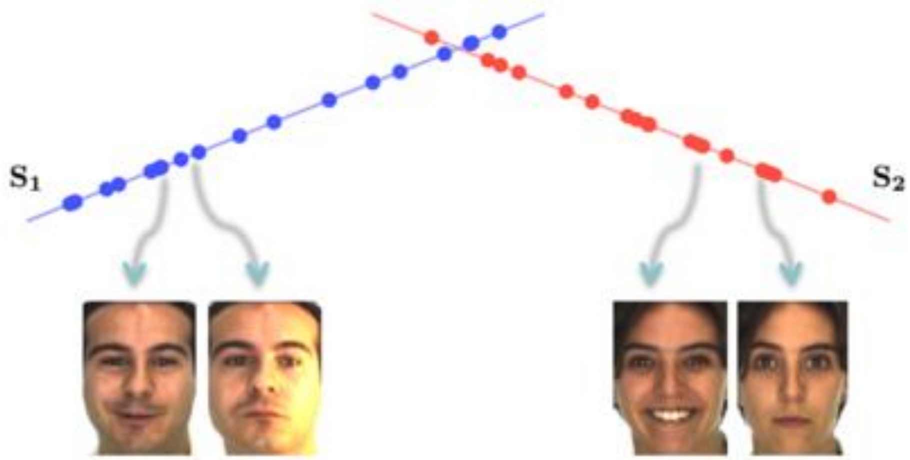
Scalable Sparse Subspace Clustering by Orthogonal Matching Pursuit

Chong You[†], Daniel P. Robinson[‡], René Vidal[†]

[†]Center for Imaging Science, Johns Hopkins University
[‡]Applied Mathematics and Statistics, Johns Hopkins University

Low-dimensional, multi-class data

- Data contains multiple classes.
- Each class lies in a low-dimensional subspace.

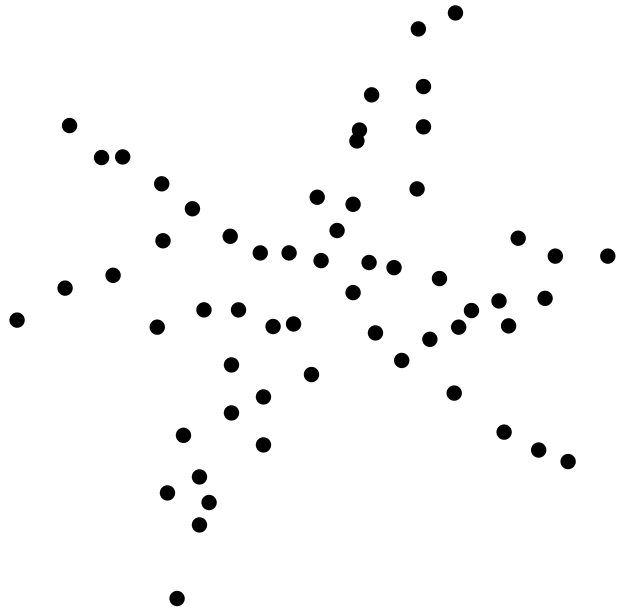


Face Recognition/Clustering

Motion Segmentation

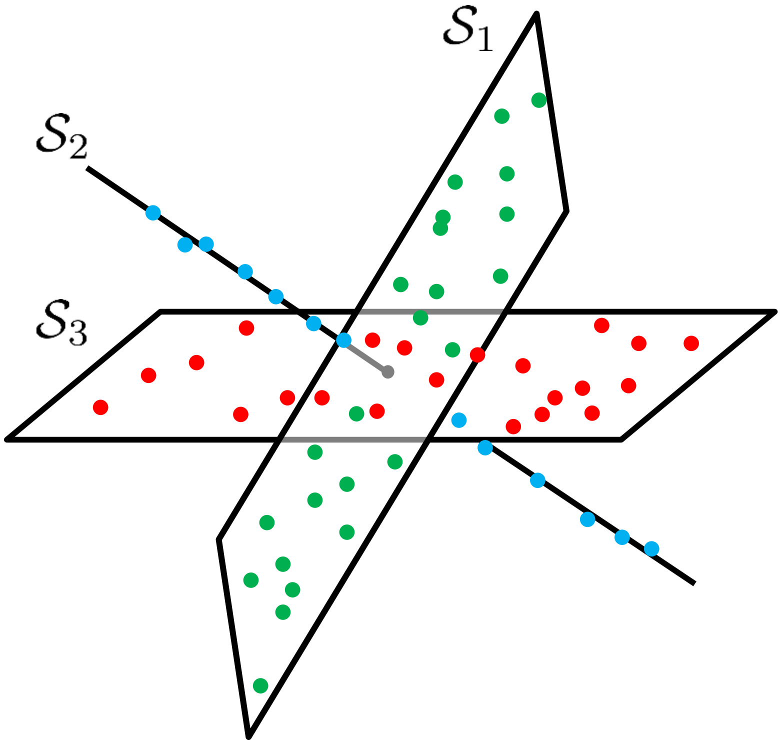
Task: subspace clustering

Given data $X = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ lying in a union of subspaces, find the segmentation

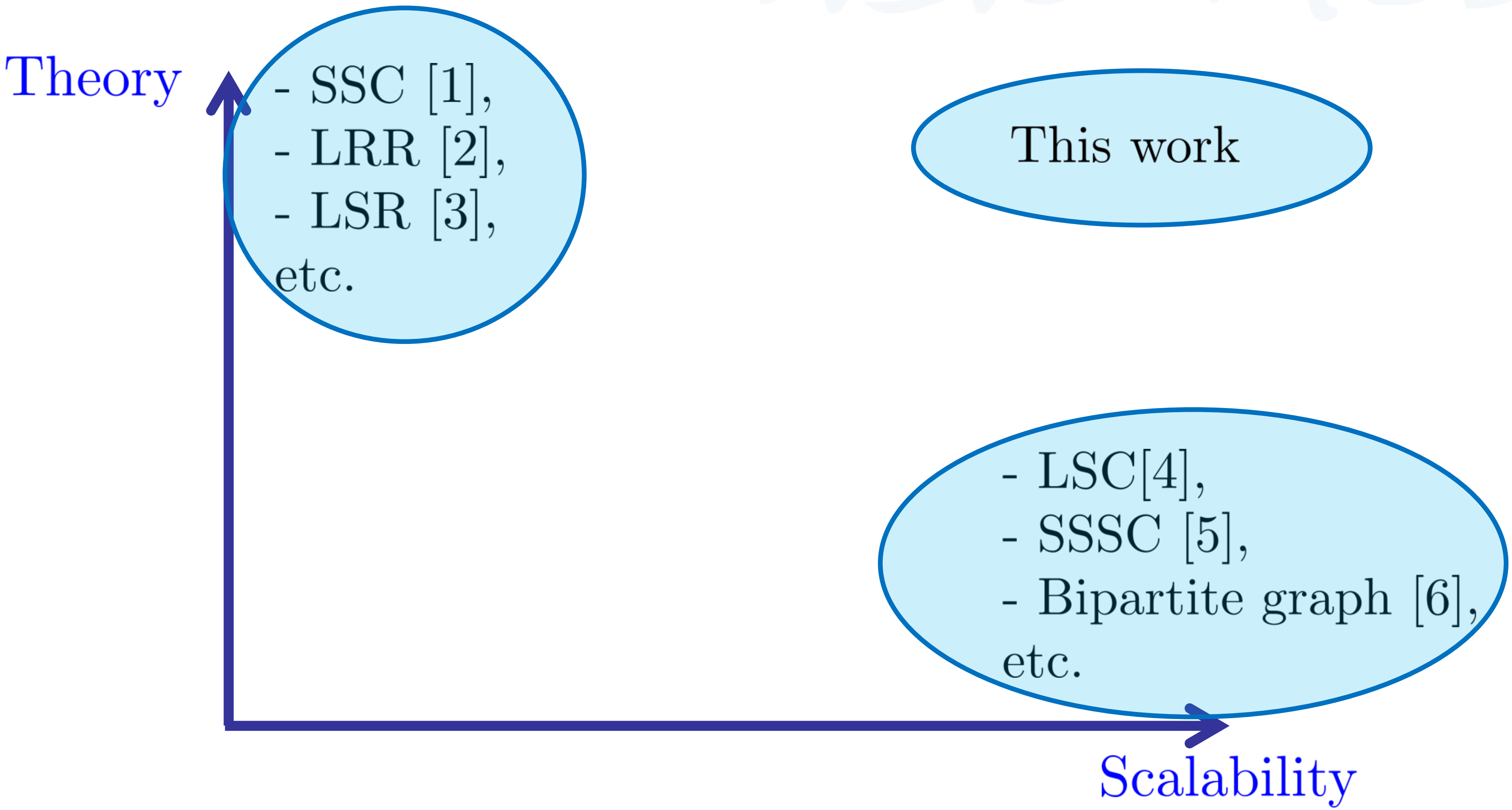


Task: subspace clustering

Given data $X = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ lying in a union of subspaces, find the segmentation

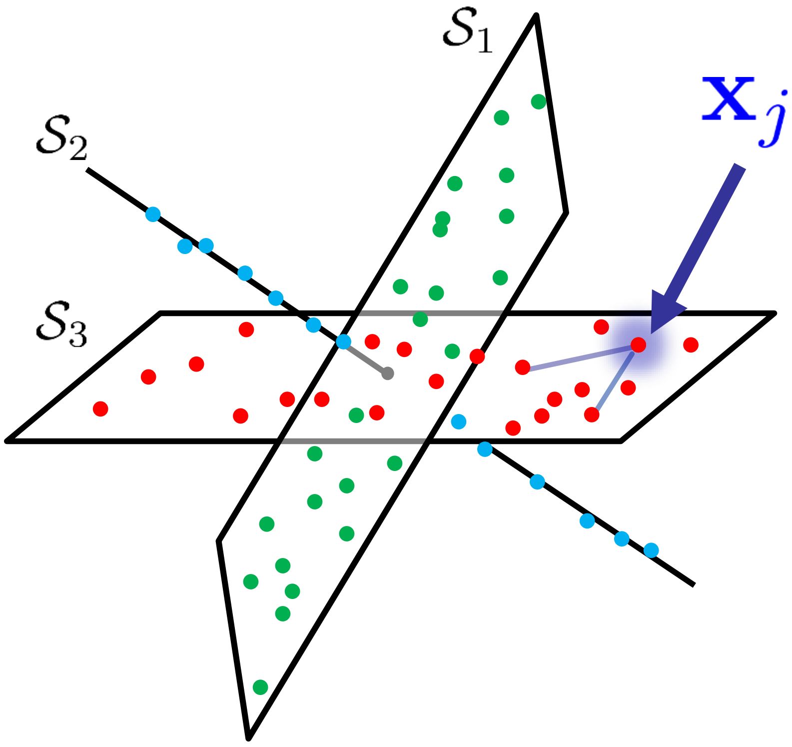


Prior work: overview



[1] E. Elhamifar and R. Vidal, Sparse Subspace Clustering, CVPR'09
[2] G. Liu, Z. Lin, Y. Yu, Robust Subspace Segmentation by Low-Rank Representation, ICML'10
[3] Lu et al,, Robust and efficient subspace segmentation via least squares regression, ECCV 2012.
[4] X. Chen and D. Cai, Large Scale Spectral Clustering with Landmark-based Representation, AAAI'11
[5] X. Peng, L. Zhang, Z. Yi, Scalable Sparse Subspace Clustering, CVPR'13
[6] A. Adler, M. Elad, Y. Hel-Or, Linear-Time Subspace Clustering via Bipartite Graph Modeling

Prior work: spectral subspace clustering



Two-step Approach

- Build data affinity
- Apply spectral clustering

Challenges

- Distance based affinity fails at the intersection of subspaces

Solution

- Compute affinity by data self-representation

Sparse Subspace Clustering (SSC) [1]:

$$\mathbf{x}_j = X\mathbf{c}_j, \quad c_{jj} = 0$$

[1] Elhamifar-Vidal, Sparse Subspace Clustering, CVPR 2009

Sparse Subspace Clustering (SSC)

$$\min_{\mathbf{c}_j} \|\mathbf{c}_j\|_0 \quad \text{s.t.} \quad \mathbf{x}_j = X\mathbf{c}_j, c_{jj} = 0$$

Basis pursuit[1] (BP)

Orthogonal matching pursuit[2] (OMP)

Method:
- Convex relaxation
- Replace $\|\mathbf{c}_j\|_0$ with $\|\mathbf{c}_j\|_1$

Properties:

✓ Guaranteed correct connections

✗ Not scalable:
solved by CVX/ADMM
tested on ≤ 640 points

Method:
- Greedy pursuit
- Choose one point at a time

Properties:

? Scalable

? Guaranteed correct connections

[1] Elhamifar-Vidal, Sparse Subspace Clustering, CVPR 2009
[2] Dyer et al, Greedy Feature Selection for Subspace Clustering, JMLR 2014

Sparse Subspace Clustering (SSC)

$$\min_{\mathbf{c}_j} \|\mathbf{c}_j\|_0 \quad \text{s.t.} \quad \mathbf{x}_j = X\mathbf{c}_j, c_{jj} = 0$$

Basis pursuit[1] (BP)

Orthogonal matching pursuit[2] (OMP)

Method:

- Convex relaxation
- Replace $\|\mathbf{c}_j\|_0$ with $\|\mathbf{c}_j\|_1$

Properties:

✓ Guaranteed correct connections

✗ Not scalable:
solved by CVX/ADMM
tested on $\leq$ 640 points

Method:

- Greedy pursuit
- Choose one point at a time

Contributions:

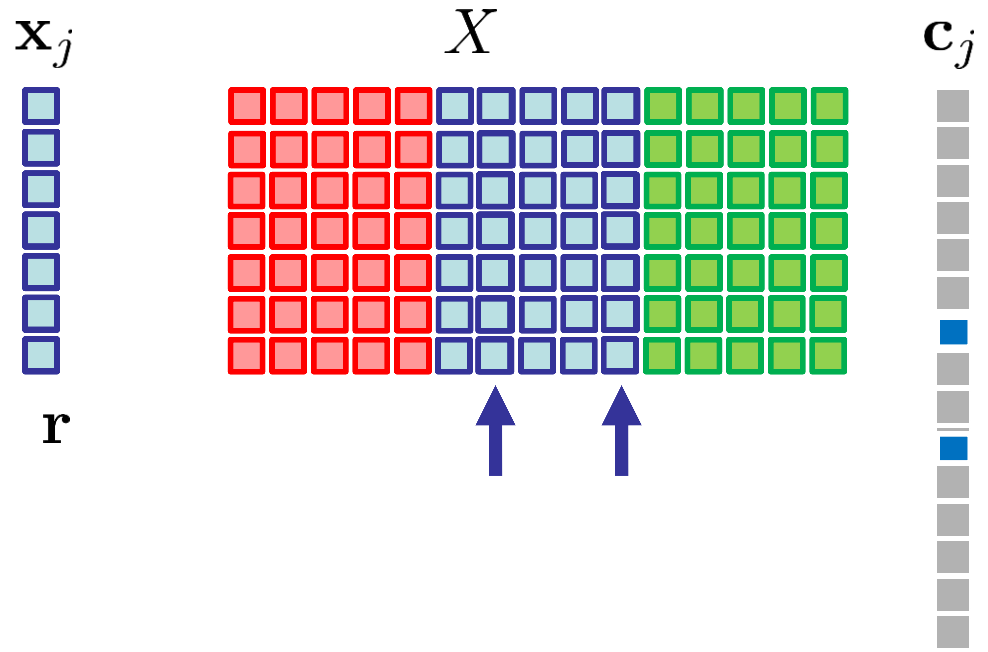
✓ Guaranteed correct connections

✓ Scalable:
tested on 100,000 points

[1] Elhamifar-Vidal, Sparse Subspace Clustering, CVPR 2009
[2] Dyer et al, Greedy Feature Selection for Subspace Clustering, JMLR 2014

SSC by orthogonal matching pursuit

Find representation $\mathbf{x}_j = X\mathbf{c}_j$ by greedy selection



What are the conditions for giving correct connections?

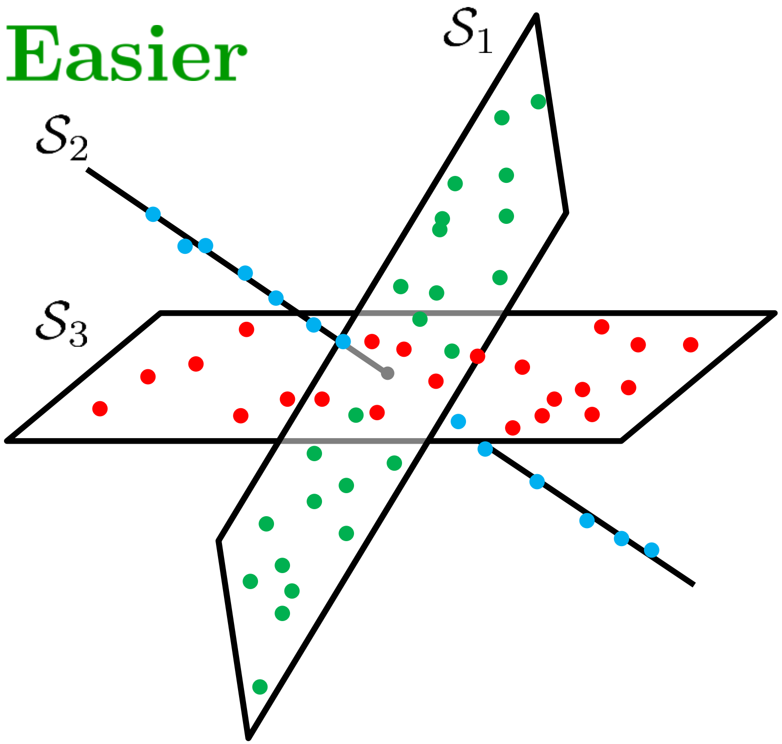
Each iteration picks a point from the same subspace

Geometric conditions for guaranteed correct connections

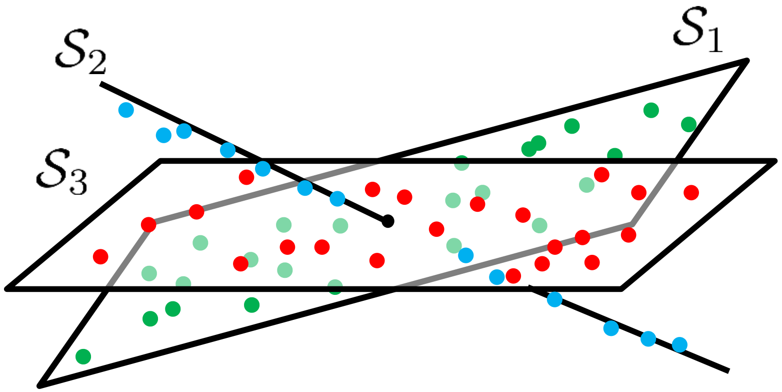
$\mu(W^\ell, X^{-\ell}) < r^\ell$

Similarity between subspaces

Easier



Harder



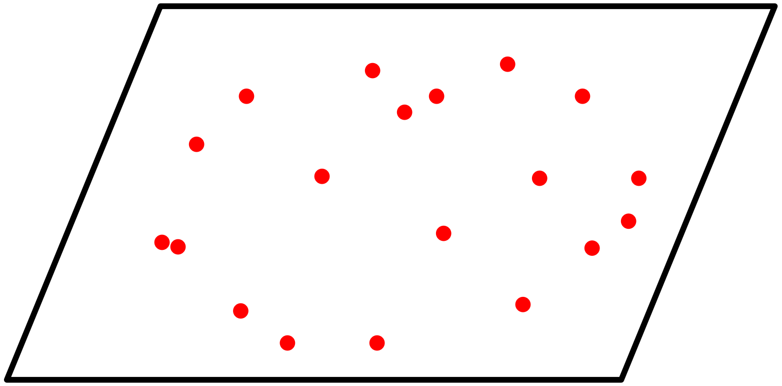
Geometric conditions for guaranteed correct connections

$$\mu(W^{\ell}, X^{-\ell}) < r^{\ell}$$

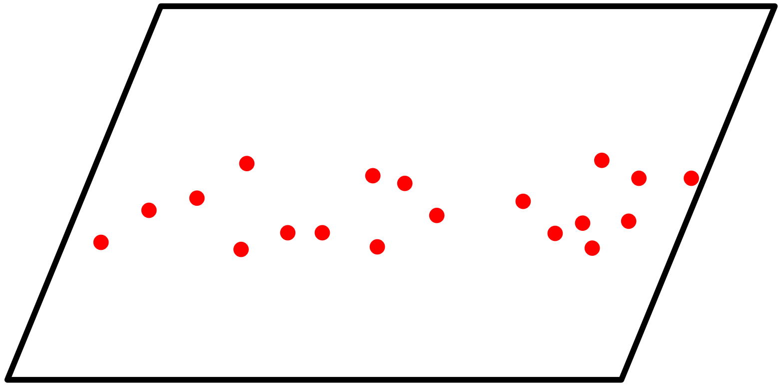
Similarity between subspaces

Distribution of points

Easier



Harder



Geometric conditions for guaranteed correct connections

$$\mu(W^{\ell}, X^{-\ell}) < r^{\ell}$$

Similarity between subspaces

Distribution of points

- Same for SSC-BP and SSC-OMP, different in definition of W^{ℓ}

? Is this condition likely to be satisfied

Guaranteed correct connections: random model

Random model:

- Draw n subspaces of dimension d in ambient dimension D
- Draw equal number of points from each subspace

Theorem:

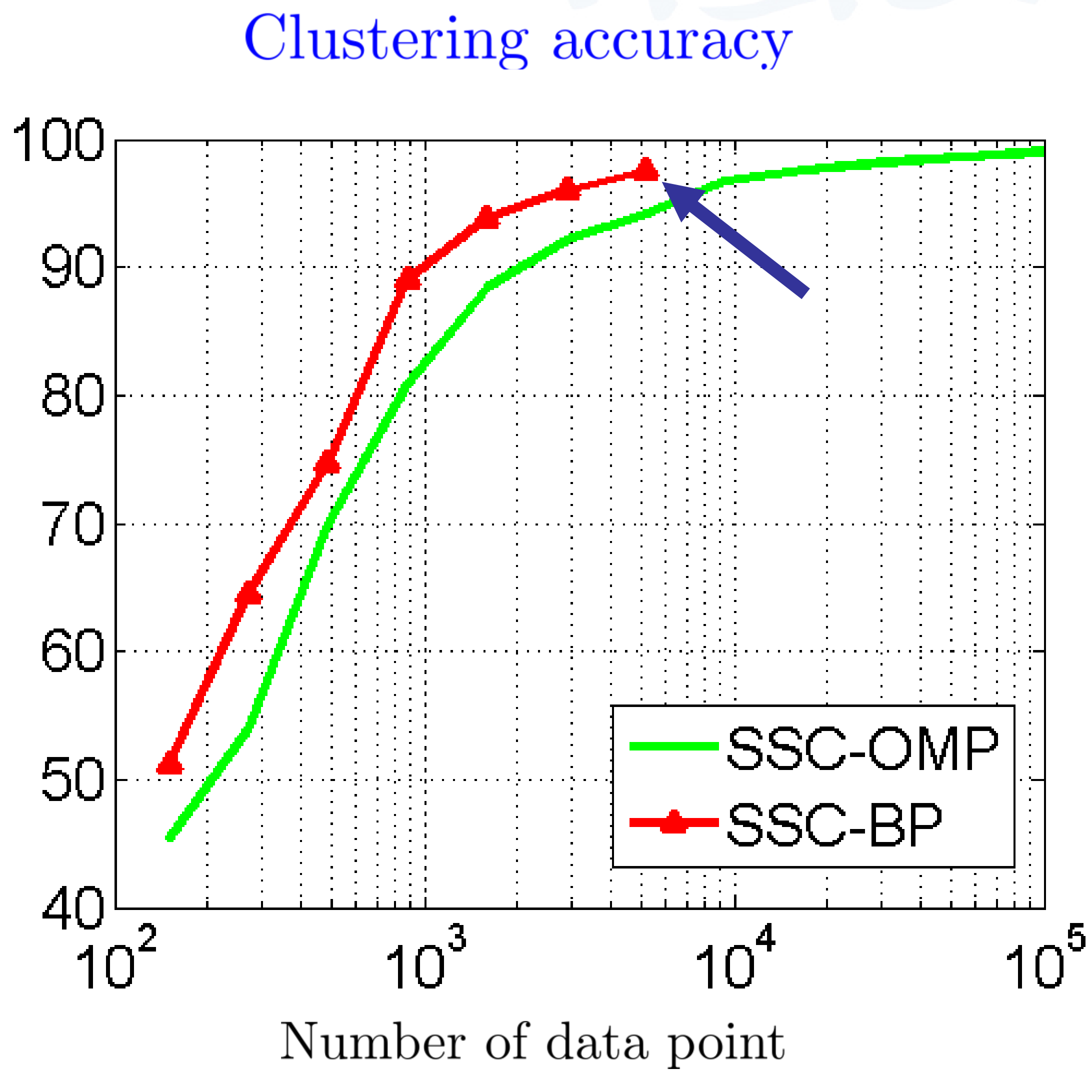
SSC-OMP is guaranteed to give correct connections with overwhelming probability if

$$\frac{d}{D} \text{ is small}$$

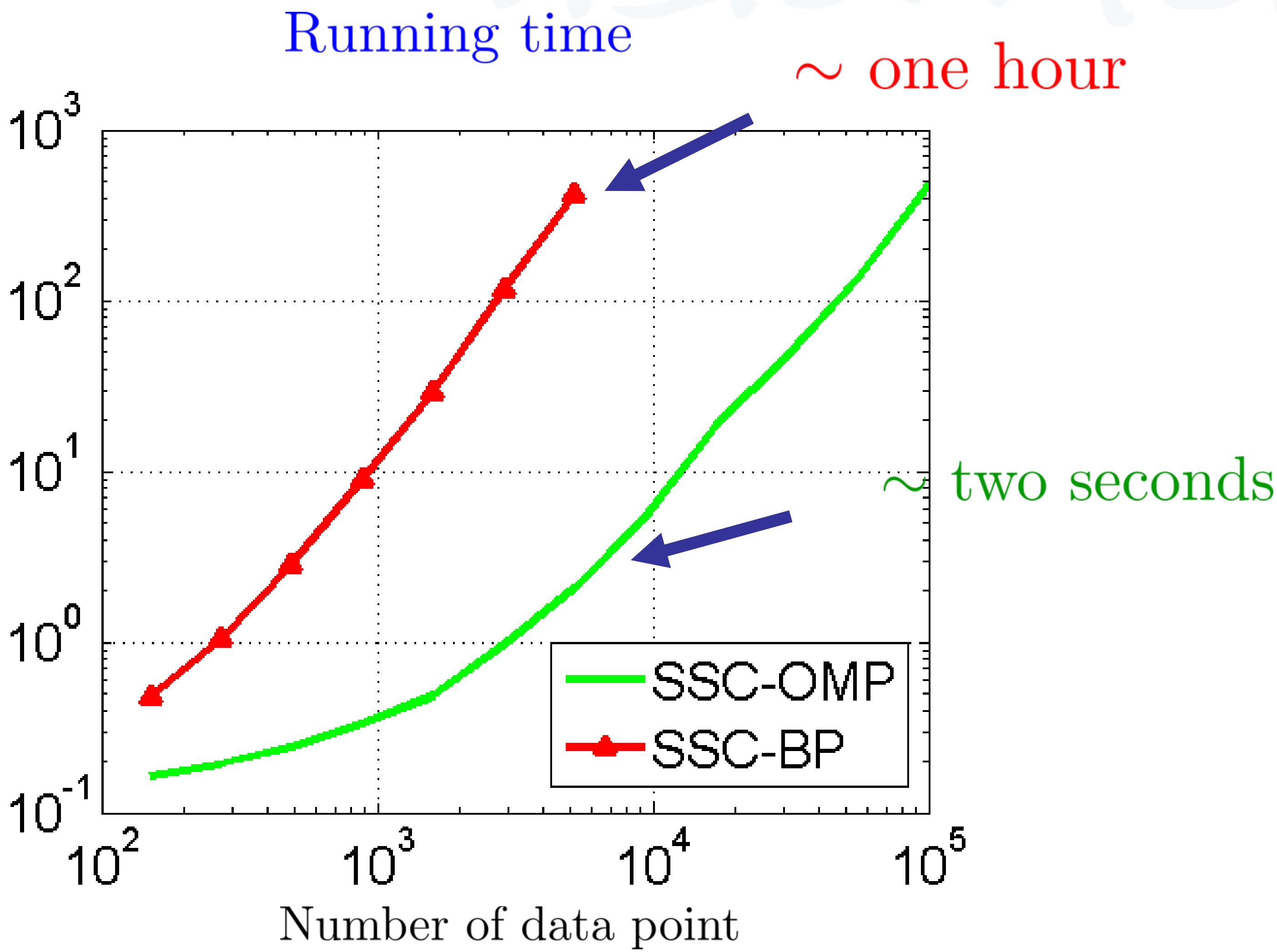
$$p(\text{SSC-BP}) - p(\text{SSC-OMP}) = O(d/N)$$

[3] Soltanolkotabi-Candes, A geometric analysis of subspace clustering with outliers

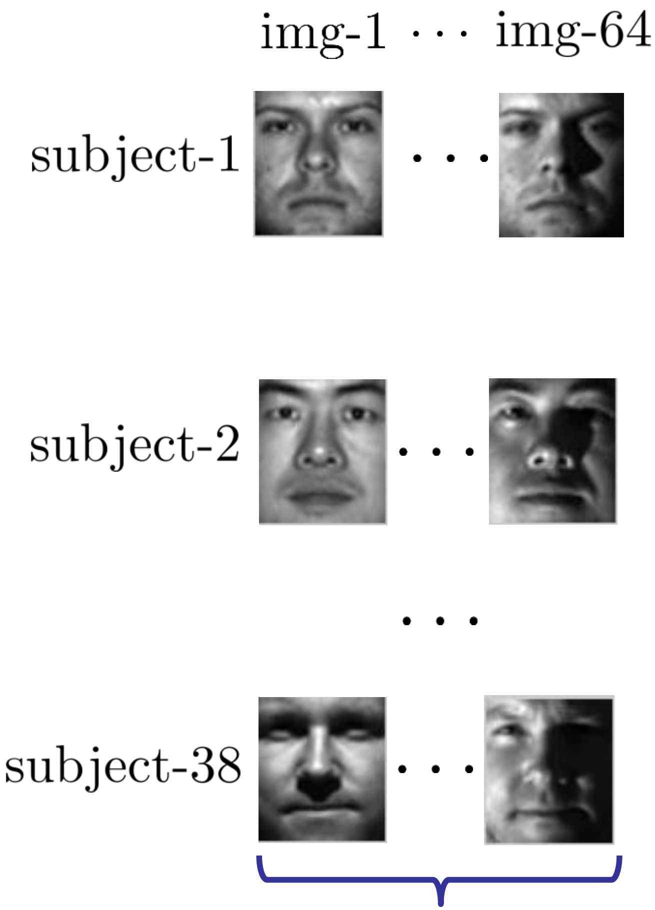
Synthetic experiments



Synthetic experiments



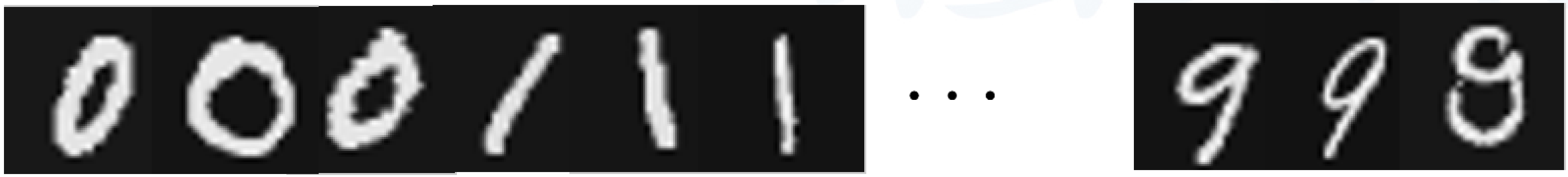
Experiment on extended Yale B



No. subjects	2	10	20	30	38
a%: average clustering accuracy					
SSC-OMP	99.21	88.43	**81.71**	**79.27**	**80.45**
SSC-BP	**99.45**	**91.85**	79.80	76.10	68.97
LSR	96.77	62.89	67.17	67.79	63.96
LRSC	94.32	66.98	66.34	67.49	66.78
SCC	78.91	NA	NA	14.15	12.80
t(sec.): running time					
SSC-OMP	0.3	1.7	4.7	9.4	**14.5**
SSC-BP	49.1	228.2	554.6	1240	1851
LSR	**0.1**	**0.8**	**3.1**	**8.3**	15.9
LRSC	1.1	1.9	6.3	14.8	26.5
SCC	50.0	NA	NA	520.3	750.7

> 100 times faster

Experiment on MNIST



No. points	500	2,000	6,000	20,000	60,000
a%: average clustering accuracy					
SSC-OMP	**85.17**	**88.99**	**90.56**	**94.21**	**94.68**
SSC-BP	83.01	85.58	85.60	-	-
LSR	75.84	78.09	79.91	-	-
LRSC	75.02	79.44	79.88	-	-
SCC	53.45	66.43	70.60	-	-
t(sec.): running time					
SSC-OMP	**1.3**	**11.7**	**71.7**	**427**	**3219**
SSC-BP	20.1	635.2	13605	-	-
LSR	1.7	42.4	327.6	-	-
LRSC	1.9	43.0	312.9	-	-
SCC	31.2	101.3	366.8	-	-

Conclusion

SSC by Orthogonal Matching Pursuit (OMP):

- ✓ stronger theoretical guarantees for correct connections
- ✓ performance validation on large databases

Acknowledgement

Funding: NSF-IIS 1447822

Vision Lab @ Johns Hopkins University
http://www.vision.jhu.edu

Thank you for your attention!

Scalable Elastic Net Subspace Clustering

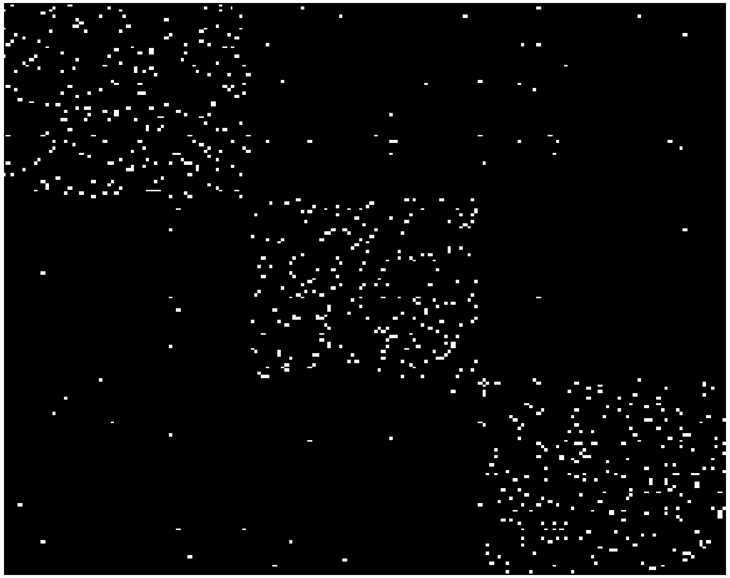
Chong You[†], Chun-Guang Li[*], Daniel P. Robinson[‡], René Vidal[†]

[†]Center for Imaging Science, Johns Hopkins University
[*]SICE, Beijing University of Posts and Telecommunications
[‡]Applied Mathematics and Statistics, Johns Hopkins University

Motivation

SSC

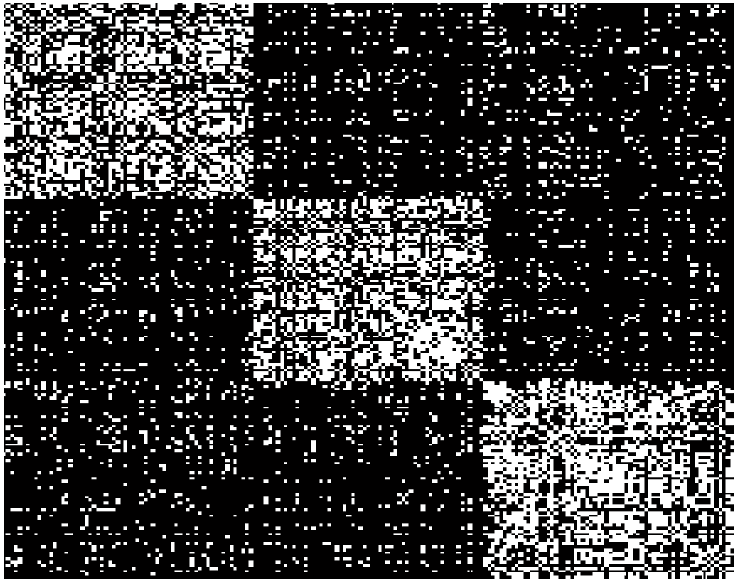
$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 + \frac{\gamma}{2}\|\mathbf{x} - X\mathbf{c}\|_2^2$$



- ✓ Few wrong connections
- ✗ Not well connected

LSR

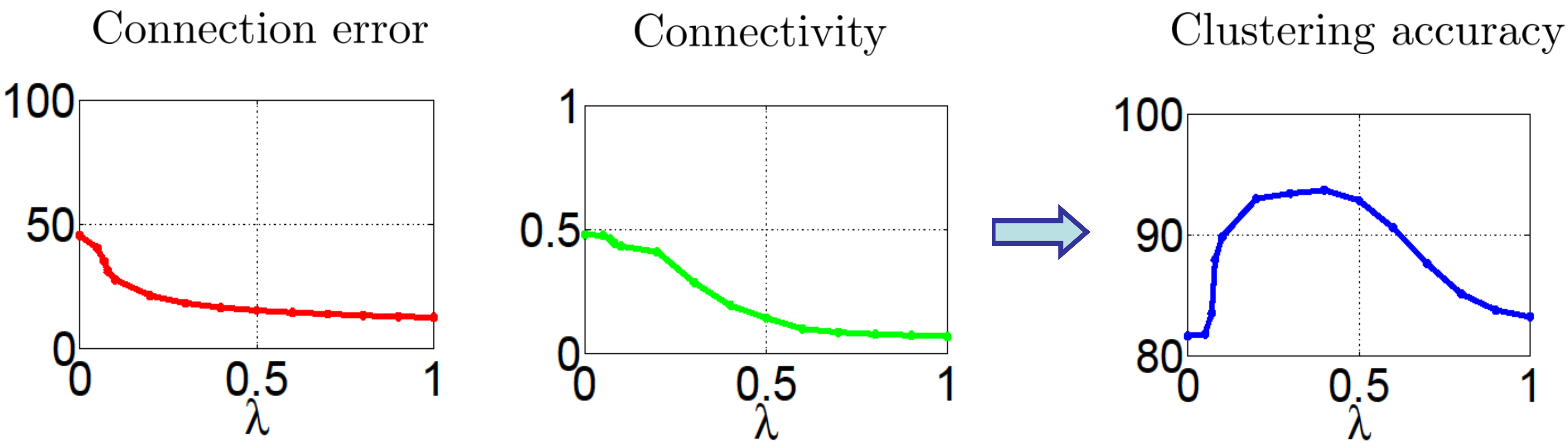
$$\min_{\mathbf{c}} \|\mathbf{c}\|_2^2 + \frac{\gamma}{2}\|\mathbf{x} - X\mathbf{c}\|_2^2$$



- ✗ Many wrong connections
- ✓ Well-connected

Elastic net Subspace Clustering (EnSC)

$$\min_{\mathbf{c}_j} \lambda\|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\gamma}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$



Contributions:

- ✔ Explain the tradeoff between correct connections and connectivity
- ✔ Prove that EnSC is guaranteed to give correct connections

Scalable Elastic net Subspace Clustering

$$\min_{\mathbf{c}_j} \lambda \|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\gamma}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$

- Prior methods
  - ADMM
  - Interior point
  - Solution path
  - Proximal gradient method
  - etc.

Scalability issue:
- Too many iterations to converge
- Access to full data matrix

Contributions:

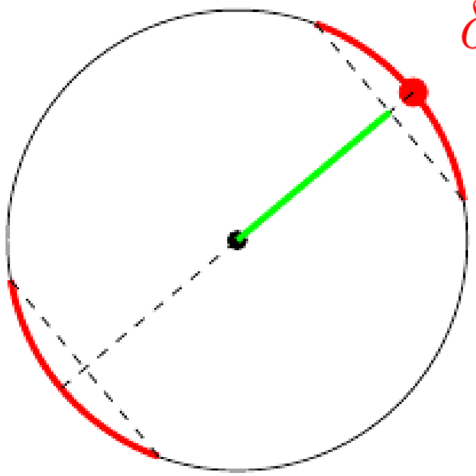
✔ Derive scalable algorithm that can handle one million data

Geometry of solution

$$\min_{\mathbf{c}_j} \lambda\|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\gamma}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$

Oracle point $\delta_j = \gamma(\mathbf{x}_j - X\mathbf{c}_j^*)$

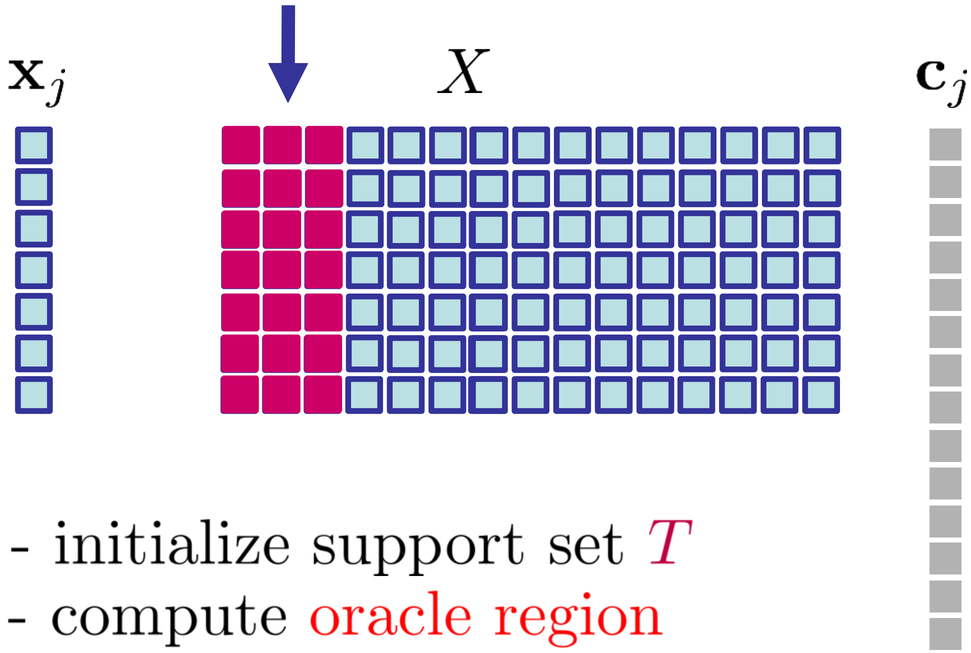
- If we know the solution $\mathbf{c}_j^*$, we can compute δ_j
- If we know δ_j, we can find the support of the solution $\mathbf{c}_j^*$



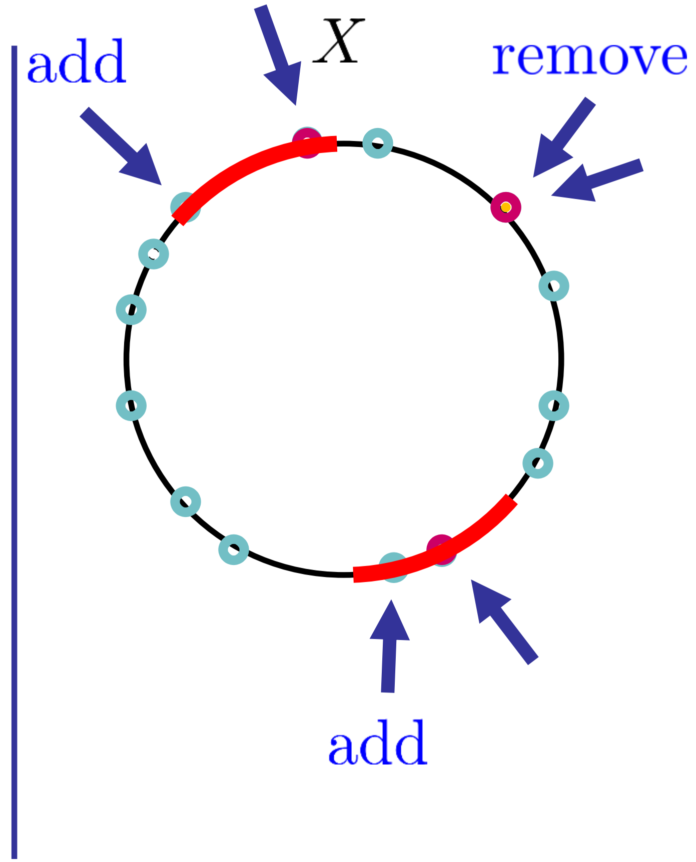
$\delta_j/\|\delta_j\|_2$

Oracle guided active set (ORGEN) algorithm

$$\min_{\mathbf{c}_j} \lambda\|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\gamma}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$



- initialize support set T
- compute oracle region



Oracle guided active set (ORGEN) algorithm

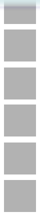
$$\min_{\mathbf{c}_j} \lambda\|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\gamma}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$

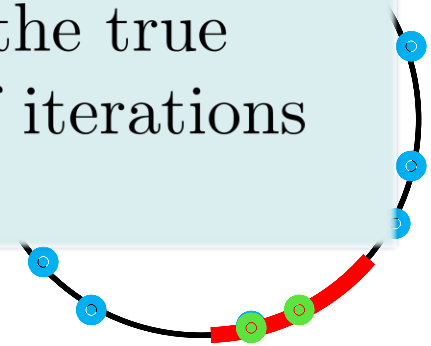
X

$\mathbf{x}_j$

Theorem:

The support set T converges to the true support set in a finite number of iterations

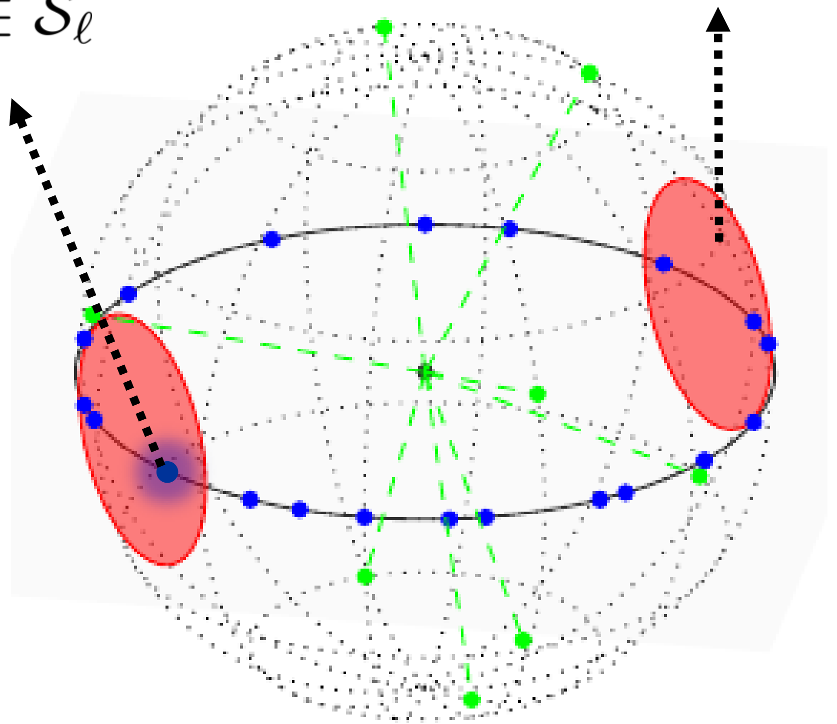




- initialize support set T
- compute oracle region
- update support set T
- repeat

Correct connections vs. connectivity

$$\min_{\mathbf{c}_j} \lambda\|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\mu}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$



- λ is large
  - $\Longrightarrow$ oracle region is small
  - $\Longrightarrow$ correct connection
- λ is small
  - $\Longrightarrow$ oracle region is large
  - $\Longrightarrow$ well-connected

Guaranteed correct connections

Theorem: (for SSC)

Theorem: (for EnSC)

Condition for guaranteed correct connections:

$$\mu(W^\ell, X^{-\ell}) < r^\ell$$

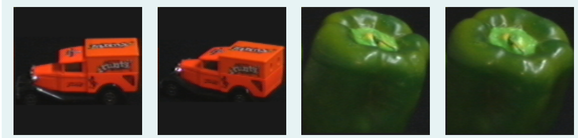
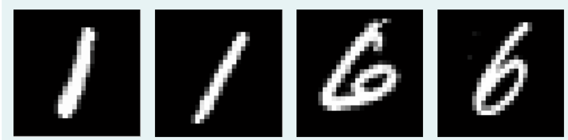
$$\mu(W^\ell, X^{-\ell}) < \frac{(r^\ell)^2}{r^\ell + \frac{1-\lambda}{\lambda}} \leq$$

- $\mu(W^\ell, X^{-\ell})$ captures similarity between subspaces
- r captures distribution of points in each subspace

(Stronger theorem is available in paper/poster)

Experiments

Test of EnSC with ORGEN on real data

database	# data	ambient dim.	# clusters	Examples
Coil-100	7,200	1024	100	
PIE	11,554	1024	68	
MNIST	70,000	500	10	
CovType	581,012	54	7	

Experiments

Our method (EnSC) achieves the best clustering accuracy

database	# data	SSC-BP	SSC-OMP	EnSC
Coil-100	7,200	57.10%	42.93%	**69.24%**
PIE	11,554	41.94%	24.06%	**52.98%**
MNIST	70,000	-	93.07%	**93.79%**
CovType	581,012	-	48.76%	**53.52%**

Experiments

Our method (EnSC) is scalable

database	# data	SSC-BP	SSC-OMP	EnSC
Coil-100	7,200	127 mins	**3 mins**	**3 mins**
PIE	11,554	412 mins	**5 mins**	13 mins
MNIST	70,000	-	**6 mins**	28 mins
CovType	581,012	-	**783 mins**	1452 mins

Conclusion

$$\min_{\mathbf{c}_j} \lambda \|\mathbf{c}_j\|_1 + \frac{1-\lambda}{2}\|\mathbf{c}_j\|_2^2 + \frac{\mu}{2}\|\mathbf{x}_j - X\mathbf{c}_j\|_2^2 \quad \text{s.t.} \quad \mathbf{c}_{jj} = 0$$

- ✓ guaranteed correct connections
- ✓ improved connectivity

} better clustering

- ✓ efficient algorithm for large scale problems

Choice of Regularization

- Prior work

Method	$f(\mathbf{c}_j)$ or $f(C)$	Correct connection[1]	Connected[2]	Scalability
SSC [1]	$\|\mathbf{c}_j\|_1$	✓		
OMP/NSN [4]	$\|\mathbf{c}_j\|_0$ (Greedy)	✓		✓
LRSSC [5]	$\|C\|_1 + \|C\|_*$	✓	✓	
LRR/LRSC [2]	$\|C\|_*$		✓	
LSR [3]	$\|\mathbf{c}_j\|_2^2$		✓	
CASS [6]	$\|X\mathrm{Diag}(\mathbf{c}_j)\|_*$		✓	
KMP [7]	$\|\mathbf{c}_j\|_k^{sp}$		✓	
EnSC (Ours)	$\|\mathbf{c}_j\|_1 + \|\mathbf{c}_j\|_2^2$	✓	✓	✓

[1]there exists theoretical guarantees for correct connection under general conditions.
[2]the solution is dense or have the grouping effect.

[1] Elhamifar-Vidal '09 '10 '13, Soltanolkotabi-Candes '12 '13, Wang-Xu '13 [2] Liu et al. '10 '13, Favaro '11, Vidal '13
[3] Lu et al. '12 [4] Dyer et al. '13, Park et al. '14, Adler et al. '15, You et al. '16 [5] Wang '13 [6] Lu et al. '13 [7] Lai et al. '14

Acknowledgement

Funding: NSF-IIS 1447822

Vision Lab @ Johns Hopkins University

http://www.vision.jhu.edu

Thank you for your attention!

Center for IMAGING SCIENCE