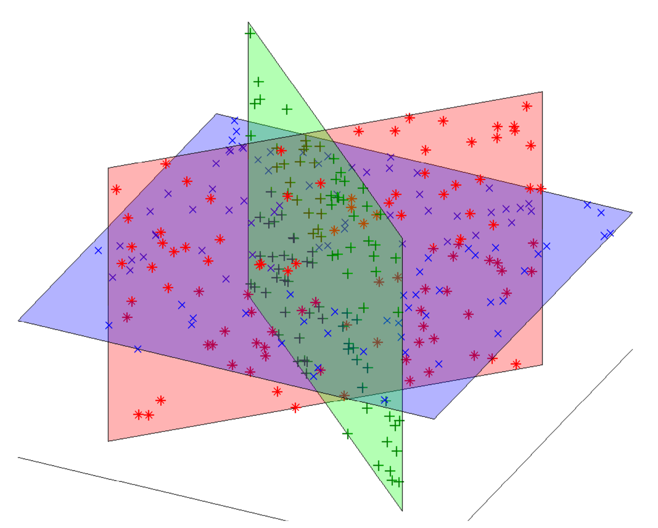
THE UNIVERSITY OF TEXAS AT AUSTIN

Greedy Subspace Clustering

Dohyung Park, Constantine Caramanis and Sujay Sanghavi

Electrical and Computer Engineering

Subspace Clustering



- Approximating data points with unions of low-dimensional subspaces
- Unions of subspaces model: Mixed datasets with latent labels, each label representing a linear subspace model.
- Task: Jointly find subspaces and cluster the points near each subspace

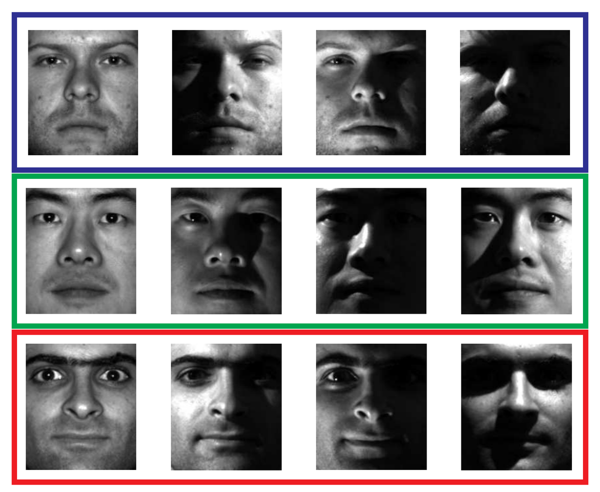
Motion Segmentation

- Feature points of a moving rigid object lie on a low-dimensional subspace

(Ref.: Hopkins155 dataset)

Face Clustering

- Face images of a person under varying illumination conditions lie on a low-dimensional subspace



(Ref.: Extended Yale B dataset)

Our contribution

Simple algorithms with

- Provable exact clustering guarantee without any subspace conditions
- Competitive performance and fast running time in practice

Related work & Comparison

Algorithms with theoretical guarantees involve two steps:
1. Neighborhood construction 2. Clustering/Subspace recovery

Algorithm	Neighborhood construction	Clustering	Sufficient conditions for: Fully random model	Semi-random model
SSC [1,2]	ℓ_1-minimization	Spectral	$\frac{d}{p} = O(\frac{\log(n/d)}{\log(nL)})$	$\max \text{aff} = O(\frac{\sqrt{\log(n/d)}}{\log(nL)})$
LRR [3]	Nuclear norm min.	Spectral	-	-
SSC-OMP [4]	OMP	Spectral	-	-
TSC [5]	Thersholding	Spectral	$\frac{d}{p} = O(\frac{1}{\log(nL)})$	$\max \text{aff} = O(\frac{1}{\log(nL)})$
LRSSC [6]	Hybrid $\ell_1 + \|\cdot\|_*$	Spectral	$\frac{d}{p} = O(\frac{1}{\log(nL)})$	-
NSN+GSR	NSN	GSR	$\frac{d}{p} = O(\frac{\log n}{\log(ndL)})$	$\max \text{aff} = O(\sqrt{\frac{\log n}{(\log dL)\cdot\log(ndL)}})$
NSN+Spectral	NSN	Spectral	$\frac{d}{p} = O(\frac{\log n}{\log(ndL)})$	-

Problem Formulation

- Subspaces are represented by their orthogonal bases $D_1, \ldots, D_L \in \mathbb{R}^{p \times d}$.
- Data points $y_1, \ldots, y_N$ have their own labels $w_1, \ldots, w_N \in [L]$ where y_i lies on subspace w_i.

$$y_i = D_{w_i} x_i, \quad x_i \sim \text{Unif}(\mathbb{S}^{d-1}), \quad \forall i \in [N].$$

- Task: Recover the labels $\{w_i : i \in [n]\}$.

Algorithms

- NSN+GSR : Run NSN to obtain neighborhood matrix $W \in \{0,1\}^{N \times N}$ and run GSR for it.
- NSN+Spectral : Run NSN to obtain W and run spectral clustering for $Z = W + W^\top$

Nearest Subspace Neighbor (NSN)

$y_i \leftarrow y_i / \|y_i\|_2, \forall i \in [N]$
for $i = 1, \ldots, N$ **do**
  $\mathcal{I}_i \leftarrow \{i\}$
  for $k = 1, \ldots, K$ **do**
    $\mathcal{U} \leftarrow \text{span}\{y_j : j \in \mathcal{I}_i\}$
    $j^* \leftarrow \arg\max_{j \in [N] \setminus \mathcal{I}_i} \|\text{Proj}_{\mathcal{U}} y_j\|_2$
    $\mathcal{I}_i \leftarrow \mathcal{I}_i \cup \{j^*\}$
  end for
  $W_{ij} \leftarrow \mathbb{I}_{j \in \mathcal{I}_i \text{ or } y_j \in \mathcal{U}}, \ \forall j \in [N]$
end for

Greedy Subspace Recovery (GSR)

$y_i \leftarrow y_i / \|y_i\|_2, \forall i \in [N]$
$\mathcal{W}_i \leftarrow \text{Top-}d\{y_j : W_{ij} = 1\}, \forall i \in [N]$
$\mathcal{I} \leftarrow [N], l \leftarrow 1$
while $\mathcal{I} \neq \emptyset$ **do**
  $i^* \leftarrow \arg\max_{i \in \mathcal{I}} \sum_{j=1}^N \mathbb{I}\{\|\text{Proj}_{\mathcal{W}_i} y_j\|_2 \geq 1 - \epsilon\}$
  $\hat{\mathcal{D}}_l \leftarrow \hat{\mathcal{W}}_{i^*}$
  $\mathcal{I} \leftarrow \mathcal{I} \setminus \{j : \|\text{Proj}_{\mathcal{W}_{i^*}} y_j\|_2 \geq 1 - \epsilon\}$
  $l \leftarrow l + 1$
end while
$\hat{w}_i \leftarrow \arg\max_{l \in [L]} \|\text{Proj}_{\hat{\mathcal{D}}_l} y_i\|_2, \ \forall i \in [N]$

Statistical Results

Fully random model - Subspaces are generated uniformly at random.

Theorem

Suppose n polynomial in d. There are constants $C_1, C_2 > 0$ such that if

$$\frac{n}{d} > C_1 \left(\log \frac{ne}{d\delta}\right)^2, \quad \frac{d}{p} < \frac{C_2 \log n}{\log(ndL\delta^{-1})},$$

then with probability $1 - \frac{3L\delta}{1-\delta}$, NSN+GSR and NSN+Spectral cluster the points exactly.

Semi-random model - Subspaces are fixed with $\max \text{aff} = \max_{i \neq j} \|D_i^\top D_j\|_F / \sqrt{d}$

Theorem

Suppose n polynomial in d. There are constants $C_1, C_2 > 0$ such that if

$$\frac{n}{d} > C_1 \left(\log \frac{ne}{d\delta}\right)^2, \quad \max \text{aff} < \sqrt{\frac{C_2 \log n}{\log(dL\delta^{-1}) \cdot \log(ndL\delta^{-1})}}.$$

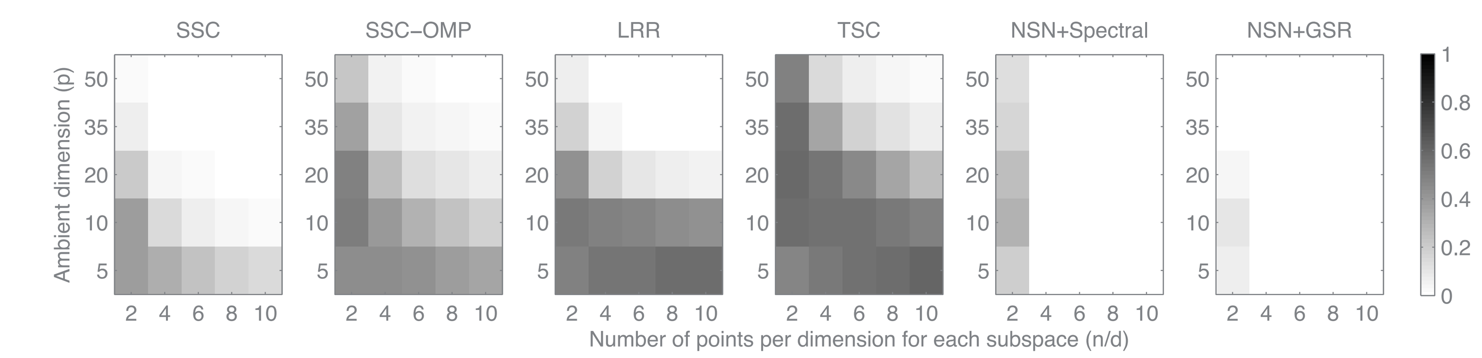
then with probability $1 - \frac{3L\delta}{1-\delta}$, NSN+GSR clusters the points exactly.

Implications

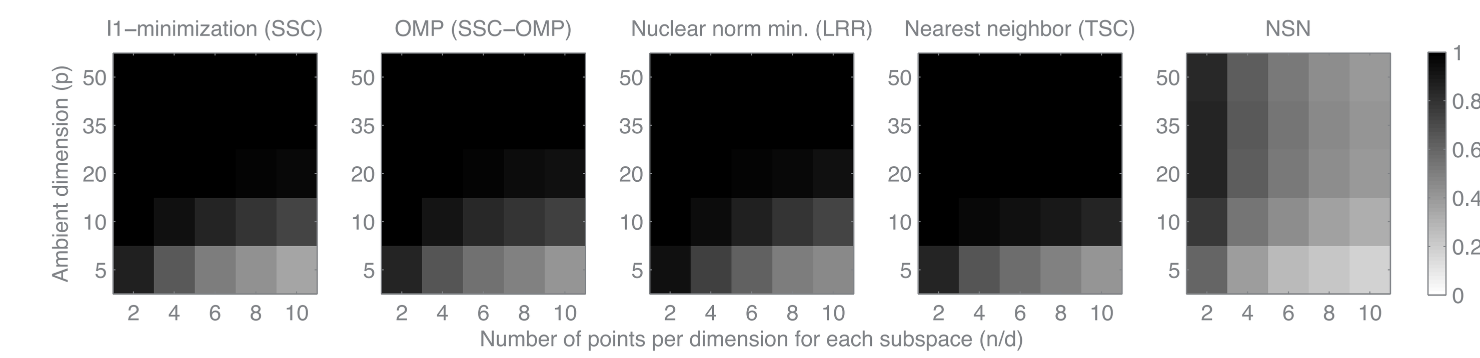
- One needs enough data points (linear in d)
- Subspaces should not be too close, but this condition improves as n grows.

Experimental Results - Synthetic Data

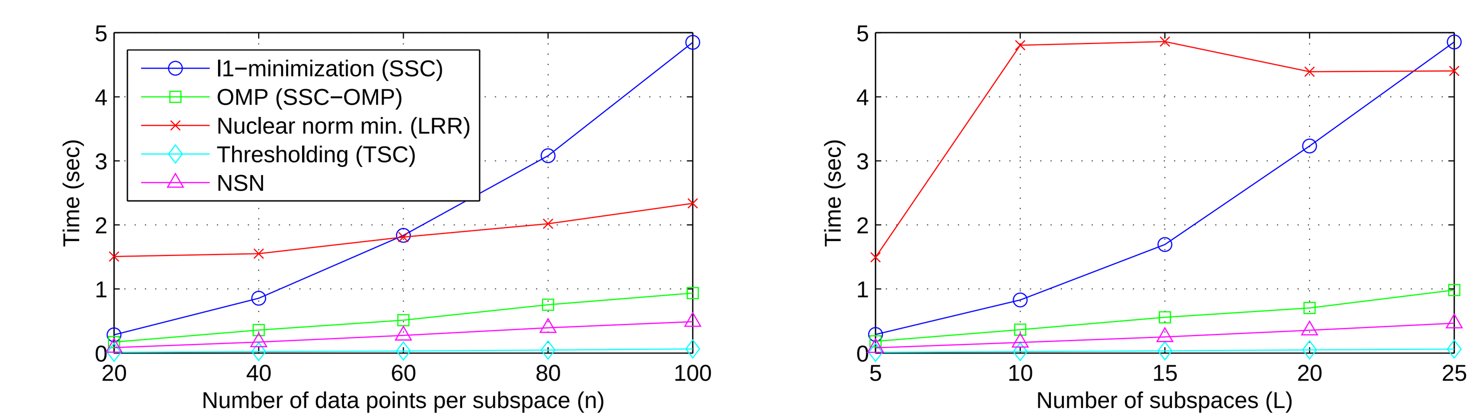
- Clustering Error in the fully random model ($d/p = 3/5$)



- Neighborhood Selection Error in the fully random model ($d/p = 3/5$)



- Average computational time for neighborhood construction



Experimental Results - Motion Segmentation

Hopkins 155 Dataset [7] - 155 video sequences of 2 or 3 motions.

L	Algorithms	K-means	K-flats	SSC	LRR	SCC	SSC-OMP	TSC	NSN+Spectral
2	Mean CE (%)	19.80	13.62	1.52	2.13	2.06	16.92	18.44	**3.62**
	Median CE (%)	17.92	10.65	0.00	0.00	0.00	12.77	16.92	**0.00**
	Avg. Time (sec)	-	0.80	3.03	3.42	1.28	0.50	0.50	**0.25**
3	Mean CE (%)	26.10	14.07	4.40	4.03	6.37	27.96	28.58	**8.28**
	Median CE (%)	20.48	14.18	0.56	1.43	0.21	30.98	29.67	**2.76**
	Avg. Time (sec)	-	1.89	5.39	4.05	2.16	0.82	1.15	**0.51**

Experimental Results - Face Clustering

Extended Yale B Dataset [8] - 64 images for each of L faces (48×42)

L	Algorithms	K-means	K-flats	SSC	SSC-OMP	TSC	NSN+Spectral
2	Mean CE (%)	45.98	37.62	1.77	4.45	11.84	**1.71**
	Median CE (%)	47.66	39.06	0.00	1.17	1.56	**0.78**
	Avg. Time (sec)	-	15.78	37.72	0.45	0.33	**0.78**
5	Mean CE (%)	73.77	55.51	4.79	8.93	11.90	**5.81**
	Median CE (%)	74.06	56.25	2.97	5.00	33.91	**4.69**
	Avg. Time (sec)	-	52.90	74.91	1.41	1.17	**5.62**
10	Mean CE (%)	82.68	62.72	9.43	15.32	39.48	**9.82**
	Median CE (%)	82.97	62.89	8.75	17.11	39.45	**9.06**
	Avg. Time (sec)	-	134.0	157.5	5.26	3.17	**14.73**

References
1. E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. IEEE TPAMI, 2013.
2. M. Soltanolkotabi and E. J. Candes. A geometric analysis of subspace clustering with outliers. The Annals of Statistics, 2012.
3. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. IEEE TPAMI, 2013.
4. E. L. Dyer, A. C. Sankaranarayanan, and R. G. Baraniuk. Greedy feature selection for subspace clustering. JMLR, 2013.
5. R. Heckel and H. Bölcskei. Robust subspace clustering via thresholding. preprint, 2014.
6. Y.-X. Wang, H. Xu, and C. Leng. Provable subspace clustering: When LRR meets SSC. In NIPS, 2013.
7. R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In CVPR, 2007.
8. A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE TPAMI, 2001.