JHU vision lab

Sparse and Low Rank Subspace Clustering

René Vidal
Center for Imaging Science
Biomedical Engineering
Johns Hopkins University

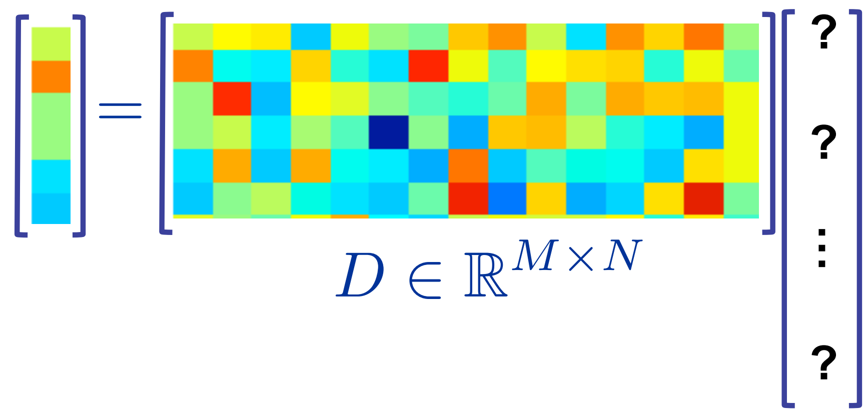
THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins

Center for IMAGING SCIENCE

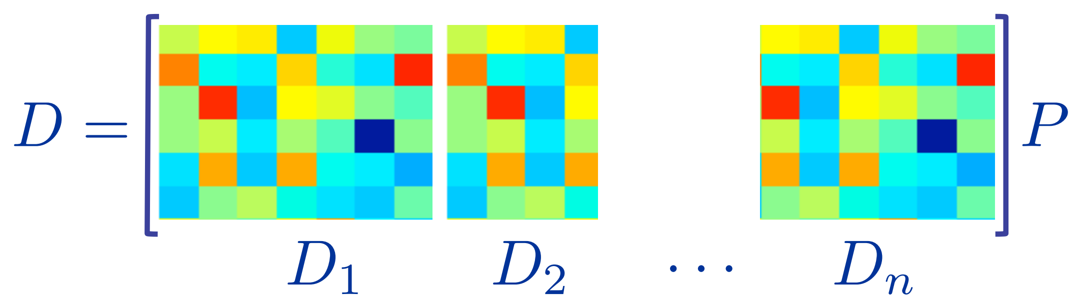
Union of Subspaces Model

- Low-rank model
  - one low-dim subspace
  - matrix completion, robust PCA

- Sparse model
  - *K* sparse signals
  - many subspaces $\binom{N}{K}$
  - equal dimensions K

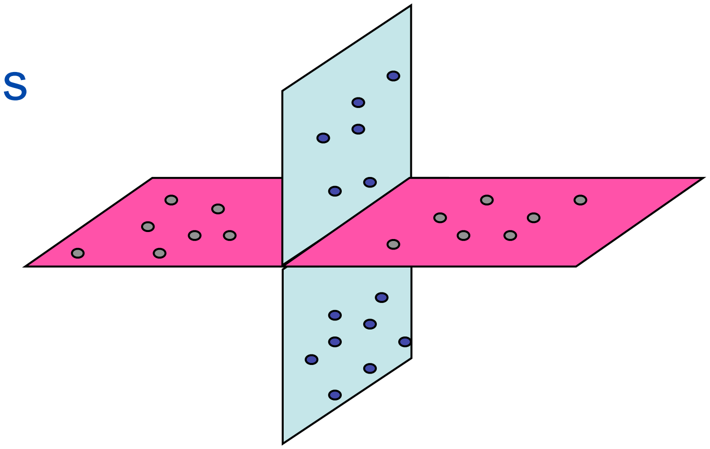


- Union of subspaces model
  - few low-dim subspaces
  - different dimensions
  - 1-block sparse signals
  - classification/clustering



Subspace Clustering Problem

- Given a set of points lying in multiple subspaces, identify
  - The number of subspaces and their dimensions
  - A basis for each subspace
  - The segmentation of the data points
- "Chicken-and-egg" problem
  - Given segmentation, estimate subspaces
  - Given subspaces, segment the data
- Challenges
  - Noise, missing entries, outliers
- Applications
  - Face/digit/speech recognition, motion/video segmentation

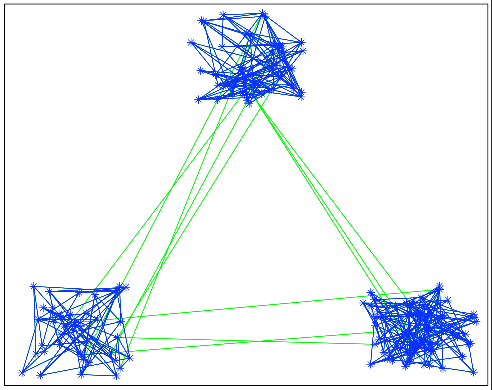


Prior Work on Subspace Clustering

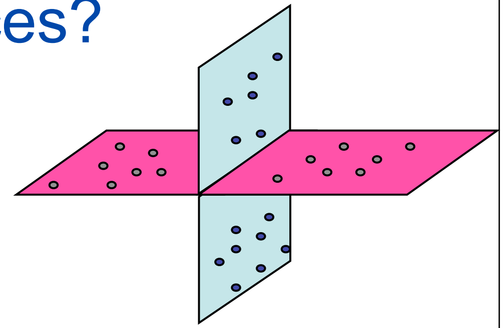
Center for IMAGING SCIENCE

Sparse and Low Rank Subspace Clustering

- Spectral clustering
  - Represent data points as nodes in graph G
  - Connect nodes i and j by edge with weight c_{ij}
  - Apply K-means to eigenvectors of the Laplacian



- How to define an affinity matrix C for subspaces?
  - Want points in the same subspace to be close
  - Want points in different subspace to be far
  - Each node connects itself to nodes in the same subspace => get a perfect block-diagonal matrix



- Data in a union of subspaces are self-expressive

$$\boldsymbol{y}_j = \sum_{i=1}^{N} c_{ij} \boldsymbol{y}_i \implies \boldsymbol{y}_j = D\boldsymbol{c}_i \implies D = DC$$

  - C is sparse
  - C is low-rank

Sparse and Low Rank Subspace Clustering

- Sparse Subspace Clustering (Elhamifar-Vidal CVPR'09, ICASSP'10)

$$\min_{C,E} \|C\|_1 + \|E\|_q \quad \text{s.t.} \quad D = DC + E \ , \ \operatorname{diag}(C) = 0$$

  - *D* is self-expressive with sparse coefficients *C*
  - Is provably correct with perfect data
  - Can handle data corrupted by noise, outliers and missing entries
  - One of the best performing algorithms for video segmentation

- Low Rank Subspace Clustering (Favaro-Vidal-Ravichandran CVPR'11)

$$\min_{A,C,E} \|C\|_* + \|E\|_q \quad \text{s.t.} \quad D = A + E \ , \ A = AC$$

  - *D* is obtained from clean self-expressive *A* with low-rank coefficients *C*
  - Is provably correct with perfect data, can handle noise and outliers
  - Important particular cases can be solved in closed form
  - Leads to a novel polynomial thresholding of the singular values

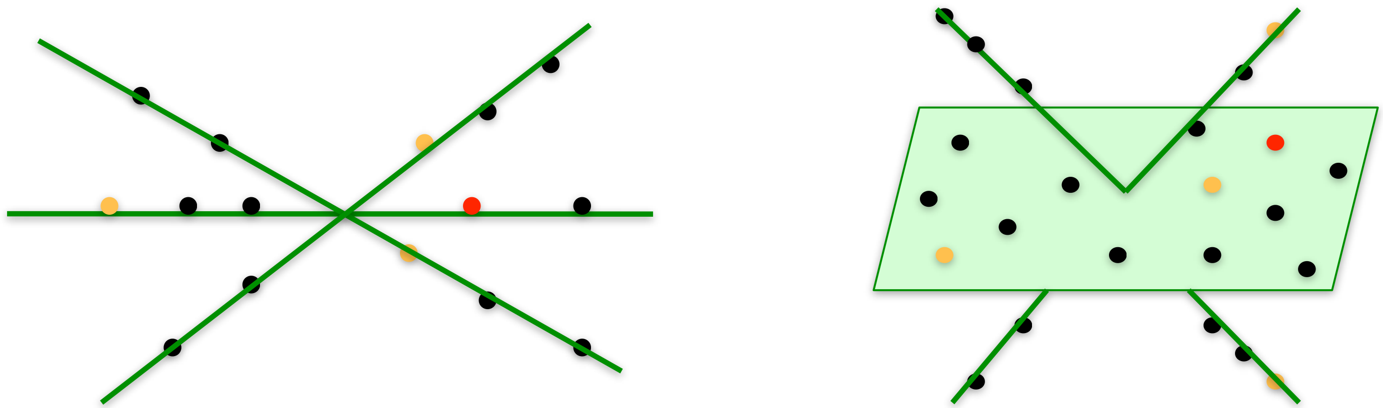
JHU vision lab

Sparse Subspace Clustering

René Vidal
Center for Imaging Science
Biomedical Engineering
Johns Hopkins University

Sparse Subspace Clustering: intuition

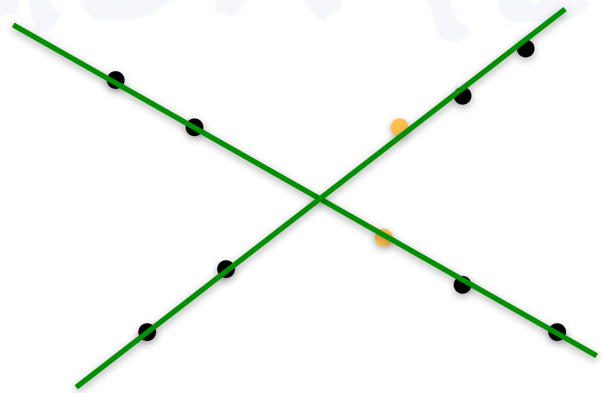
- Idea: a point $\boldsymbol{y} \in \mathbb{R}^M$ from subspace S of dimension $d \ll M$ can be written as a linear combination of d points in the same subspace $\longrightarrow$ subspace-sparse representation!



- Under what conditions on the subspaces does a sparse representation of a point come from points in the same subspace?
- Under what conditions on the subspaces can this sparse representation of a point be computed efficiently?

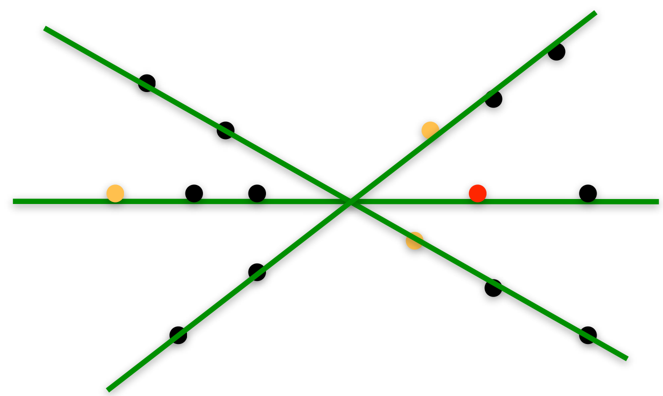
Sparse Subspace Clustering: theory

- Independent subspaces



$$\dim\left(\bigoplus_{i=1}^{n} S_i\right) = \sum_{i=1}^{n} \dim(S_i)$$

- Disjoint subspaces



$$S_i \cap S_j = \{0\}$$

- Independent implies disjoint, but disjoint does not imply independent

Sparse Subspace Clustering: theory

- *Theorem (Elhamifar & Vidal CVPR '09)*
  - *For data points drawn from a union of independent linear subspaces, a subspace-sparse representation can be found by solving the following convex program*

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{s.t.} \quad \boldsymbol{y}_i = D\boldsymbol{c}_i, \quad c_{ii} = 0$$

  - For *independent affine subspaces,*

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{s.t.} \quad \boldsymbol{y}_i = D\boldsymbol{c}_i, \quad c_{ii} = 0, \quad \text{and} \quad 1^\top \boldsymbol{c}_i = 1$$

- *Theorem (Elhamifar & Vidal ICASSP '10)*
  - *For data points drawn from a union of disjoint linear subspaces, a subspace-sparse representation can be found if the following condition on the subspace angles holds*

$$\max_{\text{rank}(\bar{\boldsymbol{Y}}_i)=d_i} \sigma_{d_i}(\bar{\boldsymbol{Y}}_i) > \sqrt{d_i} \max_{j \neq i} \cos(\theta_{ij})$$

Sparse Subspace Clustering: corrupted data

- Noise free data
  - Nonconvex problem $\min \|C\|_0 \text{ s.t. } D = DC \text{ and } \operatorname{diag}(C) = 0$
  - Convex problem $\min \|C\|_1 \text{ s.t. } D = DC \text{ and } \operatorname{diag}(C) = 0$

- Data corrupted by noise $\tilde{\boldsymbol{y}} = D\boldsymbol{c} + \boldsymbol{e}$

$$\min_{C,E} \|C\|_1 + \frac{\alpha}{2}\|E\|_F^2 \text{ s.t. } D = DC + E \text{ and } \operatorname{diag}(C) = 0$$

- Data corrupted by outliers $\tilde{\boldsymbol{y}} = D\boldsymbol{c} + \boldsymbol{e} = \begin{bmatrix} D & I \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{e} \end{bmatrix}$

$$\min_{C,E} \|C\|_1 + \|E\|_1 \text{ s.t. } D = DC + E \text{ and } \operatorname{diag}(C) = 0$$

- Data corrupted by missing entries in $I \subset \{1, \dots, M\}$
  - Form $\tilde{\boldsymbol{y}} \in \mathbb{R}^{M-|I|}$ and $\tilde{D} \in \mathbb{R}^{(M-|I|)\times N}$ by eliminating rows of $\boldsymbol{y}$ and D indexed by I, and solve the same optimization problems $\boldsymbol{y}$

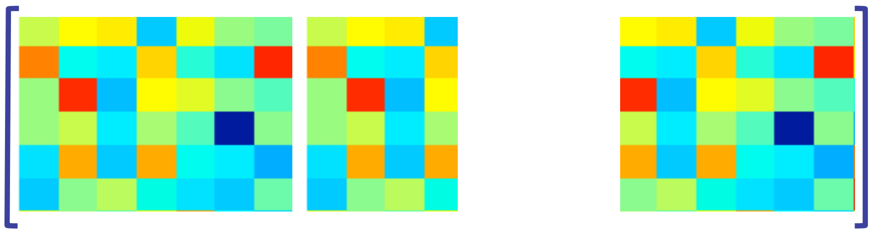
JHU vision lab

Low Rank Subspace Clustering

René Vidal
Center for Imaging Science
Biomedical Engineering
Johns Hopkins University

Low Rank Subspace Clustering (LRSC)

- Data lie in a union of low-dimensional subspaces

$D =$  $P \qquad \mathrm{rank}(D_i) = d_i \ll M$

$D_1 \quad D_2 \quad \cdots \quad D_n$

$$D_i = U_i \Sigma_i V_i^T \quad V_i V_i^T \qquad V_i \in \mathbb{R}^{N \times d_i} \quad V_i^T V_i = I_{d_i}$$

- There is a low-rank matrix of coefficients *C* $\quad \mathrm{rank}(C) \leq \sum_{i=1}^{n} d_i$

$$D = [D_1, \cdots, D_n] \begin{bmatrix} V_1 V_1^T & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & V_n V_n^T \end{bmatrix} P$$

$$D = DC$$

Low Rank Subspace Clustering (LRSC)

- Noise free data
  - Nonconvex problem $\min_{C} \ \operatorname{rank}(C) \ \text{ s.t. } \ D = DC$
  - Convex problem $\min_{C} \ \|C\|_* \ \text{ s.t. } \ D = DC$
- Theorem
  - Both optimization have a closed form solution (Liu et al. ICML'10)

$$D = U\Sigma V^T \qquad C = V_r V_r^T$$

  - If the subspaces are independent, the nonzero entries of *C* correspond to points in the same subspace (Vidal et al. IJCV'08)
- Data contaminated with noise or outliers

$$\min_{A,C,E} \ \|C\|_* + \|E\|_q \ \text{ s.t. } \ A = AC \ \text{ and } \ D = A + E$$

Low Rank Subspace Clustering (LRSC)

- The problem is nonconvex because *A* and *C* are unknown

$$\min_{A,C,E} \|C\|_* + \|E\|_q \ \text{ s.t. } \ A = AC \ \text{ and } \ D = A + E$$

- Approach
  - Case 1: Noise free and relaxed constraints
  - Case 2: Noisy data and relaxed constraints
  - Case 3: Noisy data and exact constraints
  - Case 4: Outliers

- Key contributions
  - Extend rank minimization results from one to multiple subspaces
  - Important particular cases can be solved in closed form
  - Our approach leads to a novel polynomial thresholding operator, which reduces the amount of shrinkage with respect to existing methods

Case 1: noise free data & relaxed constraint

Lemma 1 *Let* $A = U\Lambda V^T$ *be the SVD of a given matrix* A. *The optimal solution to* $\min_C \|C\|_* + \frac{\tau}{2}\|A - AC\|_F^2$ *is*

$$\widehat{C} = V_1(I - \frac{1}{\tau}\Lambda_1^{-2})V_1^T,$$

where $U = [U_1 \; U_2]$, $\Lambda = diag(\Lambda_1, \Lambda_2)$ *and* $V = [V_1 \; V_2]$ *are partitioned as* $\mathbf{I}_1 = \{i : \lambda_i > 1/\sqrt{\tau}\}$ *and* $\mathbf{I}_2 = \{i : \lambda_i \leq 1/\sqrt{\tau}\}$.

Case 2: noisy data and relaxed constraints

Lemma 2 *Let* $D = U\Sigma V^T$ *be the SVD of the data matrix* D*. The optimal solution to* $\min_{A,C} \|C\|_* + \frac{\tau}{2}\|A - AC\|_F^2 + \frac{\alpha}{2}\|D - A\|_F^2$ *is*

$$\widehat{A} = U\Lambda V^T \quad \text{and} \quad \widehat{C} = V_1(I - \frac{1}{\tau}\Lambda_1^{-2})V_1^T,$$

where each entry of Λ … *obtained from one entry of* $\Sigma = diag(\sigma_1,$ … *n to*

Polynomial thresholding

$$\widehat{A} = U\mathcal{P}_{\alpha,\tau}(\Sigma)V^T$$

$$\sigma = \psi(\lambda) = \begin{cases} \lambda + \frac{1}{\alpha\tau}\lambda^{-3} & \text{if } \lambda > 1/\sqrt{\tau} \\ \lambda + \frac{\tau}{\alpha}\lambda & \text{if } \lambda \leq 1/\sqrt{\tau} \end{cases},$$

that minimizes the cost, and the matrices $U = [U_1 \ U_2]$*,* $\Lambda = diag(\Lambda_1, \Lambda_2)$ *and* $V = [V_1 \ V_2]$ *are partitioned according to the sets* $\mathbf{I}_1 = \{i : \lambda_i > 1/\sqrt{\tau}\}$ *and* $\mathbf{I}_2 = \{i : \lambda_i \leq 1/\sqrt{\tau}\}$*.*

Case 2: polynomial thresholding operator

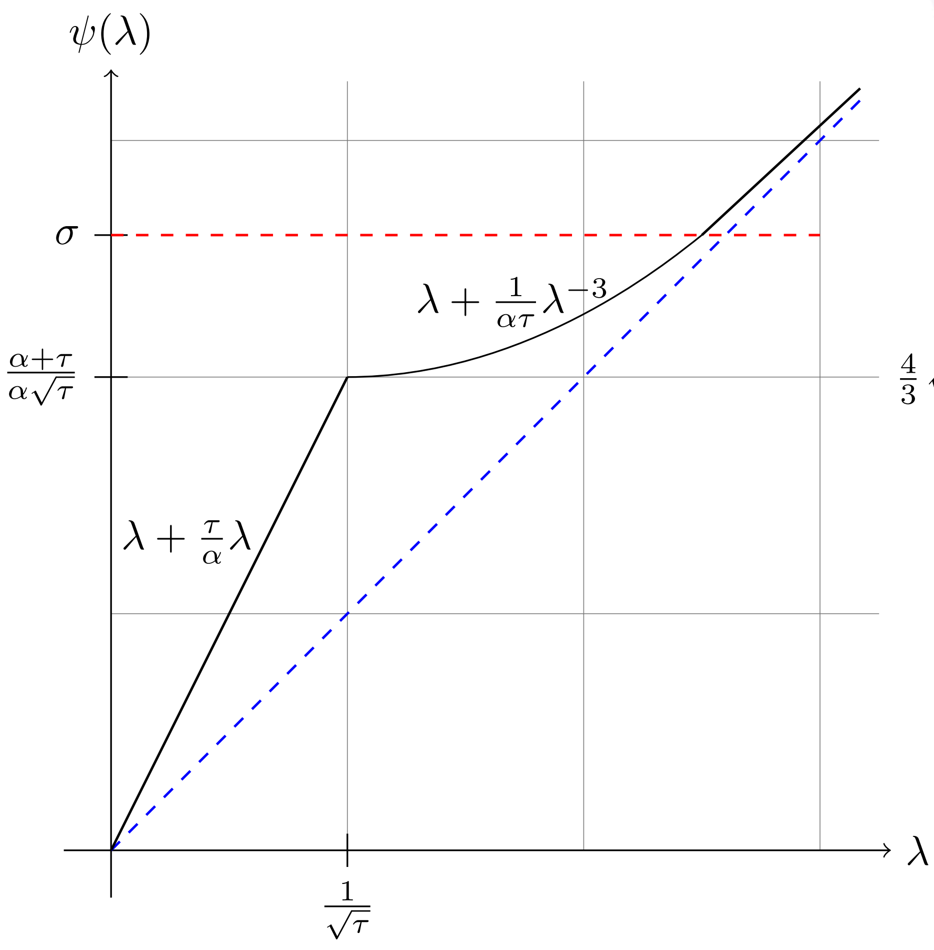


Fig. 1. Plot of $\psi(\lambda)$ when $3\tau \leq \alpha$.

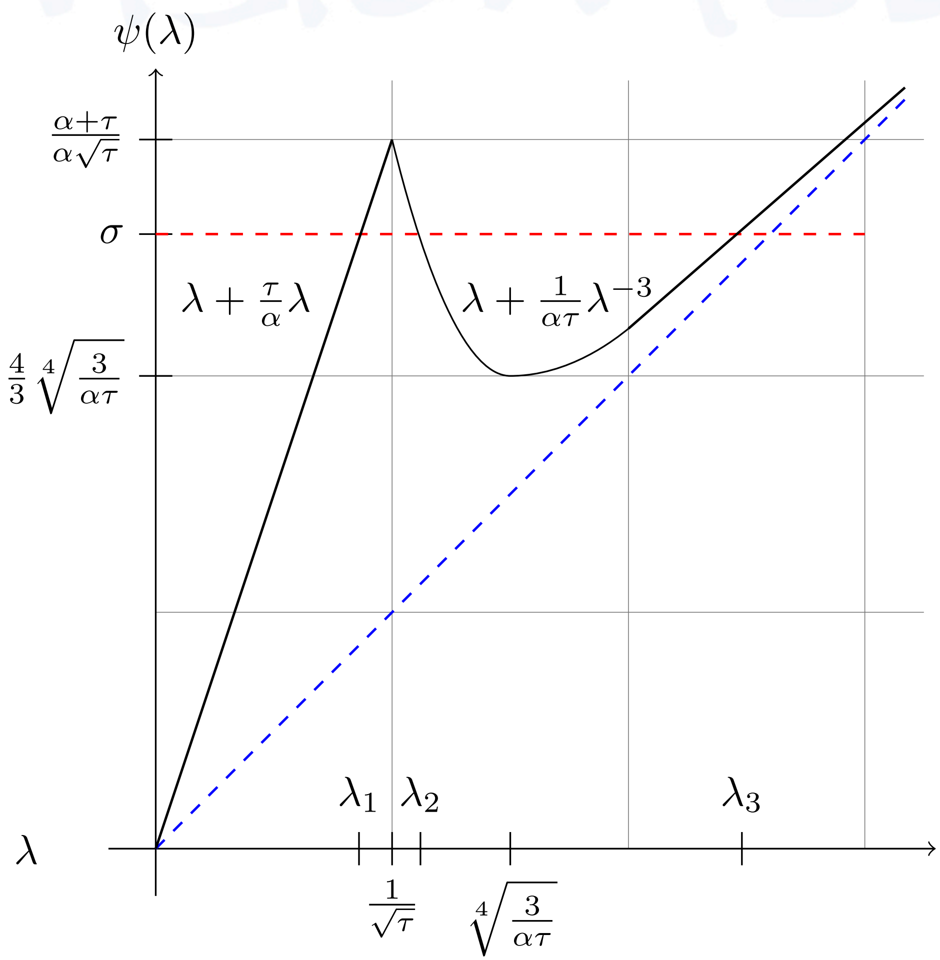


Fig. 2. Plot of $\psi(\lambda)$ when $3\tau > \alpha$.

$$\widehat{A} = U\mathcal{P}_{\alpha,\tau}(\Sigma)V^T$$

Center for IMAGING SCIENCE

Case 4: outliers

- We consider the nonconvex optimization problem

$$\min_{A,C,E} \|C\|_* + \gamma\|E\|_1 \text{ s.t. } A = AC \text{ and } D = A + E$$

- We use the Augmented Lagrangian Method (ALM)

$$\min_{A,C,E} \|C\|_* + \frac{\alpha}{2}\|D - A - E\|_F^2 + < Y, D - A - E > + \gamma\|E\|_1$$

- We obtain the iterative polynomial thresholding algorithm

$$(U,S,V) = \mathrm{svd}(D - E_k + \alpha_k^{-1}Y_k)$$

$$A_{k+1} = U\mathcal{P}_{\alpha_k,\tau}(S)V^T$$

$$E_{k+1} = \mathcal{S}_{\gamma\alpha_k^{-1}}(D - A_{k+1} + \alpha_k^{-1}Y_k)$$

$$Y_{k+1} = Y_k + \alpha_k(D - A_{k+1} - E_{k+1})$$

Polynomial thresholding

$$A_{k+1} = U\mathcal{P}_{\alpha_k,\tau}(\Sigma)V^T$$

Shrinkage thresholding

$$A_{k+1} = U\mathcal{S}_{\frac{1}{\alpha_k}}(\Sigma)V^T$$

Center for IMAGING SCIENCE

JHU vision lab

Applications in Computer Vision

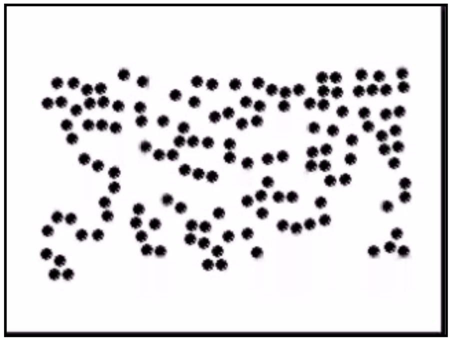
René Vidal
Center for Imaging Science
Biomedical Engineering
Johns Hopkins University

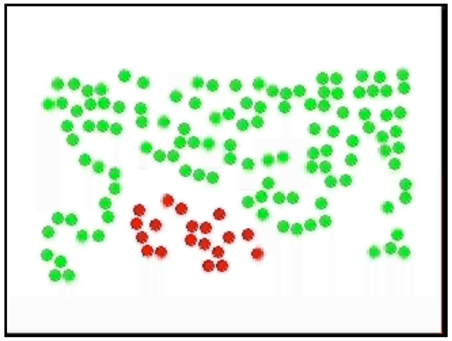
THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins

Center for IMAGING SCIENCE

Segmentation of Dynamic Scenes

- Motion segmentation problem
  - Input: multiple images of a scene with multiple rigid-body motions
  - Output: number of motions, motion model parameters, segmentation





- Motion of a rigid-body lives in 4D linear subspace (Boult and Brown '91, Tomasi and Kanade '92)
  - P = #points
  - F = #frames

$$
\mathtt{W} = \mathtt{M} \quad \mathtt{S}^T
$$

$$
\underbrace{\begin{bmatrix} \boldsymbol{x}_{11} \cdots \boldsymbol{x}_{1P} \\ \vdots \qquad \vdots \\ \boldsymbol{x}_{F1} \cdots \boldsymbol{x}_{FP} \end{bmatrix}}_{2F \times P} = \underbrace{\begin{bmatrix} \mathtt{A}_1 \\ \vdots \\ \mathtt{A}_F \end{bmatrix}}_{2F \times 4} \underbrace{\begin{bmatrix} \boldsymbol{X}_1 \cdots \boldsymbol{X}_P \end{bmatrix}}_{4 \times P}
$$

Vidal et al., ECCV02, IJCV06; Vidal, Ma and Sastry CVPR03, PAMI05; Vidal and Sastry CVPR03; Vidal and Ma ECCV04, JMIV06; Vidal and Hartley, CVPR04; Tron and Vidal, CVPR07; Li et al. CVPR07; Goh and Vidal CVPR07; Vidal and Hartley, PAMI08; Vidal, Tron and Hartley IJCV08; Rao et al. CVPR 08, PAMI 09; Elhamifar and Vidal, CVPR 09

Results on the Hopkins 155 database

- 2 motions, 120 sequences, 266 points, 30 frames

	GPCA	LLMC	LSA	RANSAC	MSL	SCC	ALC	SSC-B	SSC-N
Checkerboard	6.09	3.96	2.57	6.52	4.46	1.30	1.55	**0.83**	1.12
Traffic	1.41	3.53	5.43	2.55	2.23	1.07	1.59	0.23	**0.02**
Articulated	2.88	6.48	4.10	7.25	7.23	3.68	10.70	1.63	**0.62**
All	4.59	4.08	3.45	5.56	4.14	1.46	2.40	**0.75**	0.82

- 3 motions, 35 sequences, 398 points, 29 frames

	GPCA	LLMC	LSA	RANSAC	MSL	SCC	ALC	SSC-B	SSC-N
Checkerboard	31.95	8.48	5.80	25.78	10.38	5.68	5.20	4.49	**2.97**
Traffic	19.83	6.04	25.07	12.83	1.80	2.35	7.75	0.61	**0.58**
Articulated	16.85	9.38	7.25	21.38	2.71	10.94	21.08	1.60	**1.42**
All	28.66	8.04	9.73	22.94	8.23	5.31	6.69	3.55	**2.45**

- All

	GPCA	LLMC	LSA	RANSAC	MSL	SCC	ALC	SSC	LRR	LCSR
All	10.34	4.97	4.94	9.76	5.03	2.33	3.37	1.24	3.16	3.28

Results with missing entries & outliers

- Misclassifications rates on 12 motion sequences with missing data

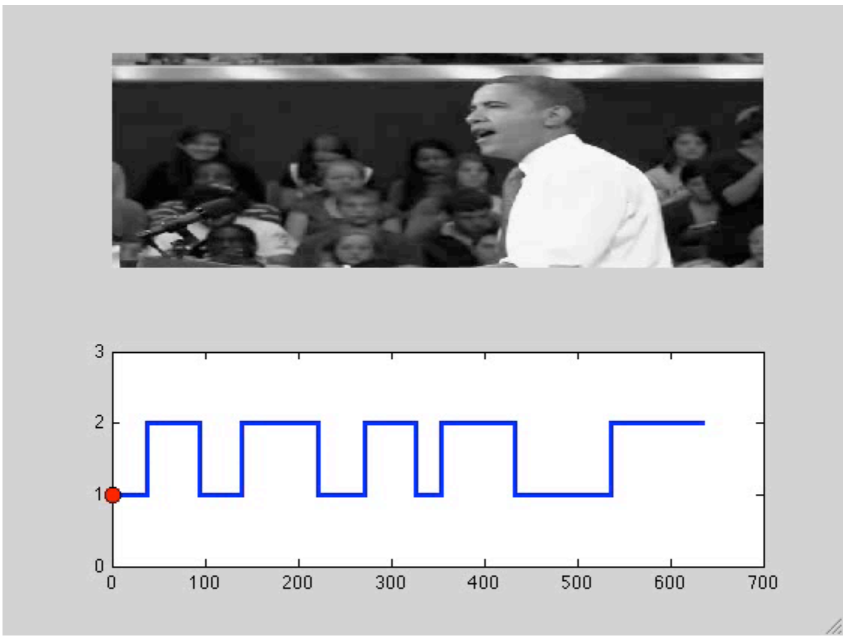
Method	PF+ ALC_5	PF+ALC_{sp}	ℓ^1+ALC_5	ℓ^1+ALC_{sp}	SSC-N
Average	1.89%	10.81%	3.81%	1.28%	**0.13%**
Median	0.39%	7.85%	0.17%	1.07%	0.00%

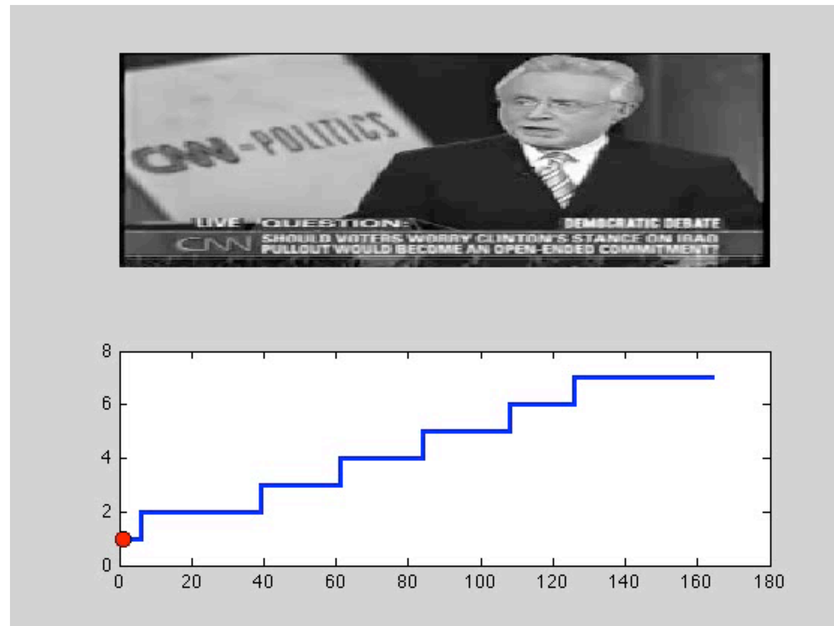
- Misclassifications rates on 12 motion sequences with corrupted data

Method	$\ell^1 + \text{ALC}_5$	$\ell^1 + \text{ALC}_{\text{sp}}$	SSC	LRSC
Average	4.15%	3.02%	**1.05%**	1.22%

Temporal Video Segmentation by SSC

- Model each video segment as a low-dimensional subspace
- Segment the video into multiple segments

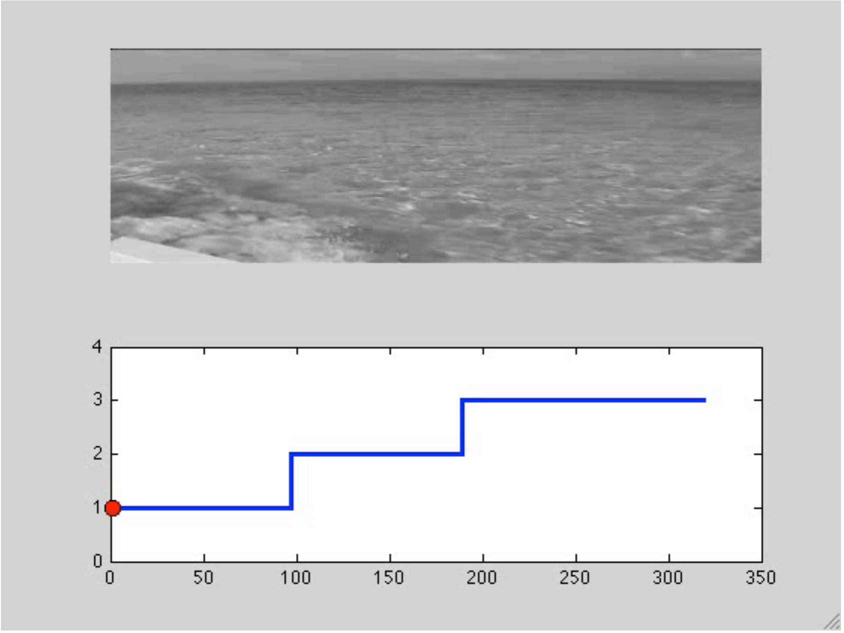


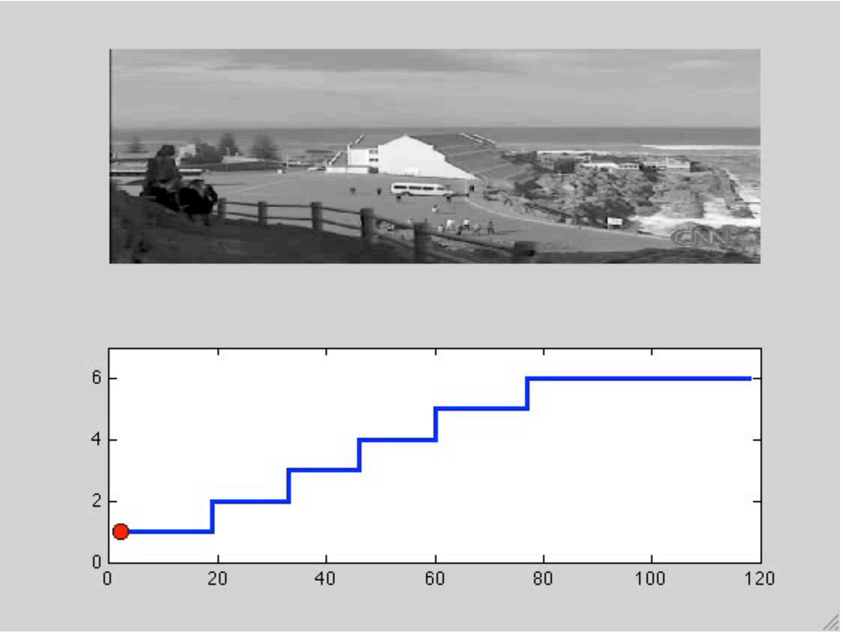


- Advantages
  - SSC easily detects sharp transitions in the video
  - SSC can handle camera motion and scene variations

Temporal Video Segmentation by SSC

- Model each video segment as a low-dimensional subspace
- Segment the video into multiple segments





- Advantages
  - SSC easily detects sharp transitions in the video
  - SSC can handle camera motion and scene variations

Conclusions

- Many problems in image processing and computer vision can be posed as multi-subspace clustering problems
  - Spatial and temporal video segmentation
  - Face clustering under varying illumination
  - Dynamic texture segmentation
- These problems can be solved using
  - Sparse Subspace Clustering (SSC): algorithm based on sparse representation theory and spectral clustering
  - Low Rank Subspace Clustering (LRSC): algorithm based on rank minimization and spectral clustering
- Future work
  - Extending SSC to nonlinear manifolds

Acknowledgements

- Collaborators
  - Paolo Favaro, HWU
  - Richard Hartley, ANU
  - Yi Ma, UIUC & Microsoft Asia
  - Shankar Rao, UIUC
- Students
  - Ehsan Elhamifar
  - Avinash Ravichandran
  - Roberto Tron
- Grants
  - Sloan Research Fellowship
  - ONR Young Investigator Award
  - ONR N00014-09-10084
  - ONR N00014-05-10836
  - NSF CAREER 0447739
  - NSF 0941463
  - NSF 0931805
  - NSF 0941362
  - NSF 0809101
  - NSF 0509101
  - ARL Robotics-CTA
  - JHU APL-934652
  - NIH RO1 HL082729
  - JHU APL-934652

Center for IMAGING SCIENCE