Low-Rank Subspace Clustering

Zhouchen Lin[1] & Jiashi Feng[2]

[1]Peking University

[2]National University of Singapore

June 26

Outline

- Representative Models
  - Low-Rank Representation (LRR) Based Subspace Clustering
  - Variants of LRR
- Analysis on LRR
  - Closed-form Solutions in Noiseless Case
  - Exact Recoverability - Deterministic Analysis
  - Exact Recoverability - Probabilistic Analysis
- Applications
- Block-Diagonal Structure in Subspace Clustering

Outline

- Representative Models
  - Low-Rank Representation (LRR) Based Subspace Clustering
  - Variants of LRR
- Analysis on LRR
  - Closed-form Solutions in Noiseless Case
  - Exact Recoverability - Deterministic Analysis
  - Exact Recoverability - Probabilistic Analysis
- Applications
- Block-Diagonal Structure in Subspace Clustering

Low-Rank Representation (LRR) Based Subspace Clustering

- Low-Rank Representation:

Enforces Correlation among Representations

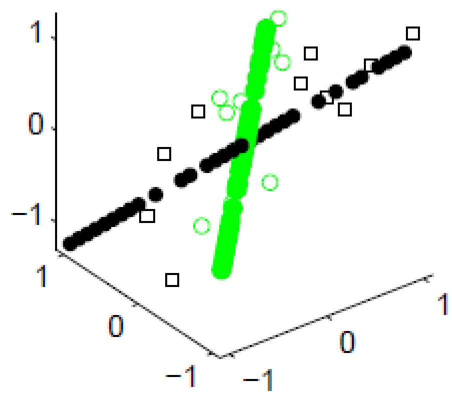
Column Sparsity

NP-Hard!

$$\min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda\|\mathbf{E}\|_{2,0}, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{E},$$

where $\|\mathbf{E}\|_{2,0} = \#\{i|\|\mathbf{E}_i\|_2 \neq 0\}$ is called the $\ell_{2,0}$-pseudonorm.

(a) Corrupted Data



Liu et al. *Robust Subspace Segmentation by Low-Rank Representation*, ICML 2010.
Liu et al., *Robust Recovery of Subspace Structures by Low-Rank Representation*, IEEE T. PAMI 2013.

Convex Relation

$$\min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{2,0}, \quad s.t. \quad \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{E}.$$

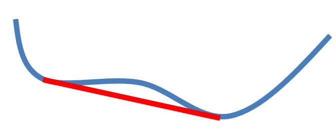
Convex Envelope

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{E}.$$

Nuclear Norm: sum of singular values

$$\boxed{||\mathbf{E}||_{2,1} = \sum_i ||\mathbf{E}_i||_2}$$

- Convex Envelope: The largest **convex** function upper bounded by the give function



- **Theorem:** The convex envelope of the rank function on the unit ball of matrix spectral norm is the nuclear norm

$$B_2(1) = \{\mathbf{X} | \|\mathbf{X}\|_2 \leq 1\}$$

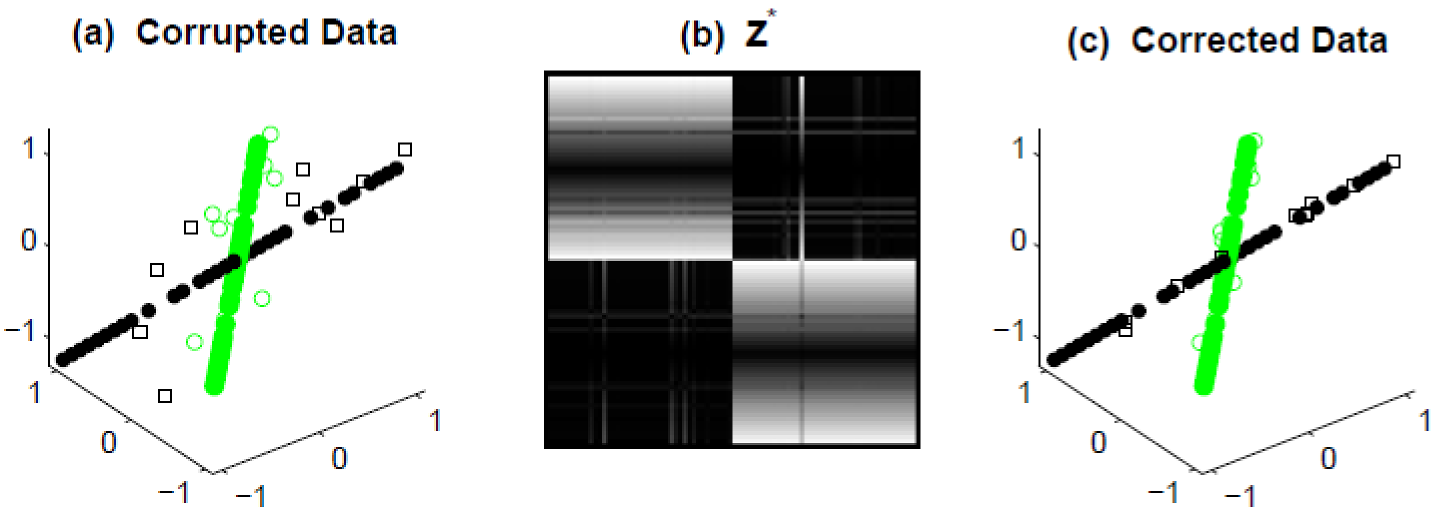
M. Fazel. *Matrix Rank Minimization with Applications*. Ph.D. Thesis, 2002.

Subspace Clustering

- Construct a graph

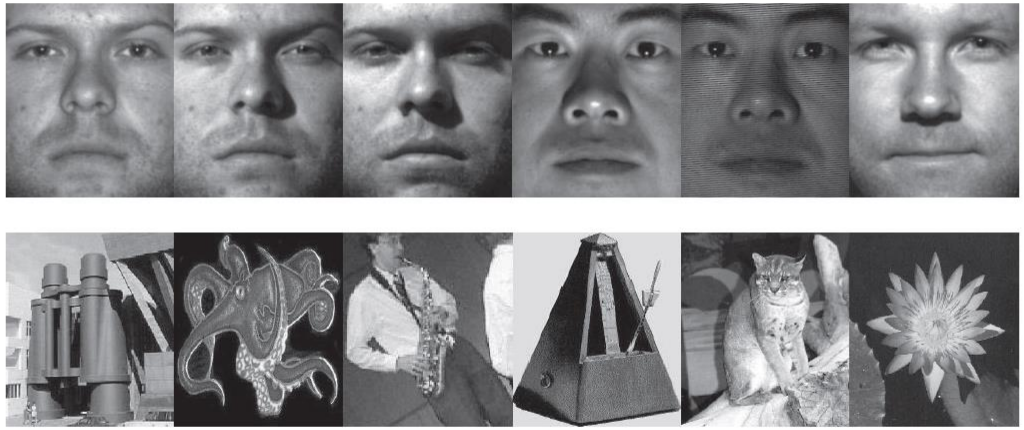
$$\mathbf{W} = (|\mathbf{Z}^*| + |(\mathbf{Z}^*)^T|)/2$$

- Normalized cut on the graph



Liu et al. *Robust Subspace Segmentation by Low-Rank Representation*, ICML 2010.
Liu et al., *Robust Recovery of Subspace Structures by Low-Rank Representation*, IEEE T. PAMI 2013.

Experimental Validation



Inliers: faces from Extended Yale B

Outliers: nonfaces from Caltech101

Examples of images in the combined Yale-Caltech dataset.

Model	PCA	RPCA_1	$\mathrm{RPCA}_{2,1}$	$\mathrm{SSC}_{2,1}$	LRR
ACC (%)	77.15	82.97	83.72	73.17	**86.13**
AUC	0.9653	0.9819	0.9863	0.9239	**0.9927**

Table 1: Segmentation Accuracy (ACC) and AUC Comparison on the Yale-Caltech Data Set.

Liu et al., *Robust Recovery of Subspace Structures by Low-Rank Representation,* IEEE T. PAMI 2013.

Variants of LRR

- Robust LRR (R-LRR)
  - Using a clean dictionary

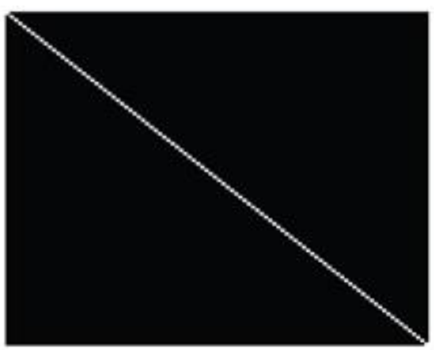
$$\min_{\mathbf{A},\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{D} = \mathbf{A} + \mathbf{E}, \mathbf{A} = \mathbf{A}\mathbf{Z}.$$

P. Favaro et al. *A closed form solution to robust subspace estimation and clustering*. CVPR 2011.
S. Wei and Z. Lin, *Analysis and improvement of low rank representation for subspace segmentation*, arXiv preprint arXiv:1107.1561.

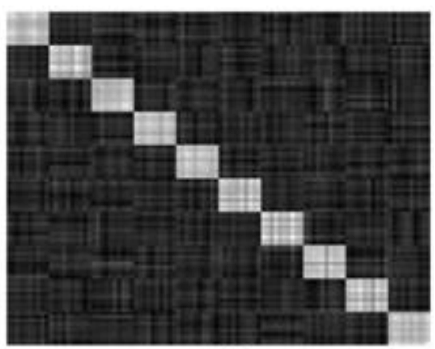
Variants of LRR

- Fix-Rank Representation
  - Addressing insufficient data

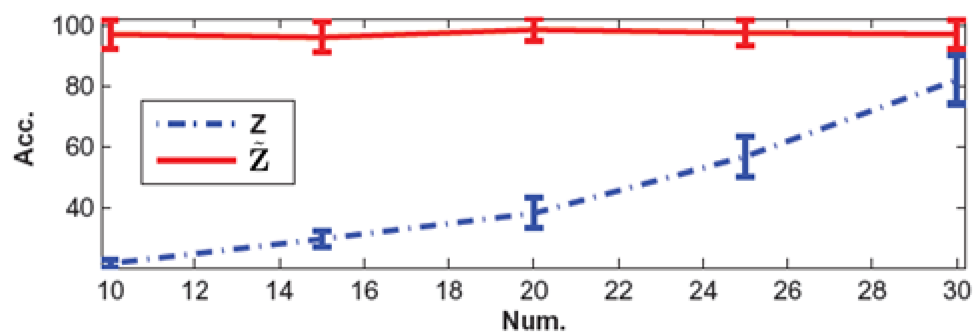
$$\min_{\mathbf{Z},\tilde{\mathbf{Z}},\mathbf{E}} \|\mathbf{Z}-\tilde{\mathbf{Z}}\|_F^2 + \lambda||\mathbf{E}||_{2,1}, \quad s.t. \quad \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{E}, \operatorname{rank}(\tilde{\mathbf{Z}}) \leq r.$$



(a) $\mathbf{Z}$



(b) $\tilde{\mathbf{Z}}$



Liu et al., *Fixed-Rank Representation for Unsupervised Visual Learning*, CVPR 2012.

Variants of LRR

- Latent LRR (LatLRR)
  - Addressing insufficient data

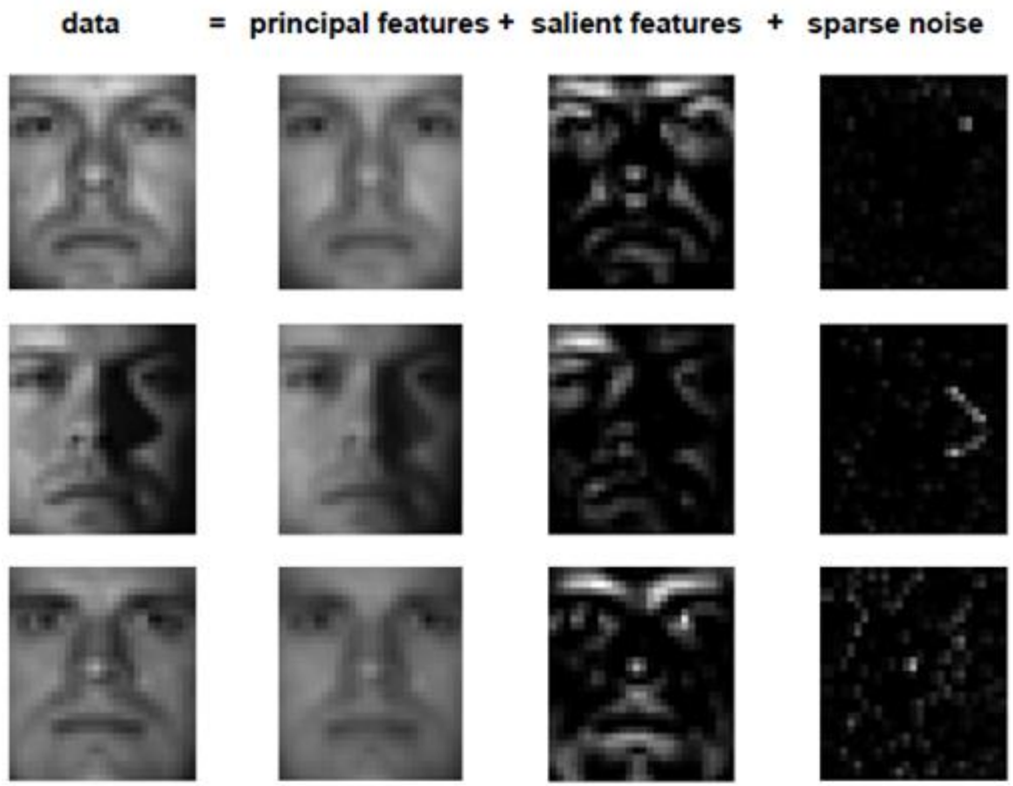
Consider unobserved samples

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*,$$
$$s.t. \quad \mathbf{D}_O = [\mathbf{D}_O, \mathbf{D}_H]\mathbf{Z}.$$

$$\min_{\mathbf{Z},\mathbf{L}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_*,$$
$$s.t. \quad \mathbf{D}_O = \mathbf{D}_O\mathbf{Z} + \mathbf{L}\mathbf{D}_O.$$

$$\min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda\|\mathbf{E}\|_1,$$
$$s.t. \quad \mathbf{D}_O = \mathbf{D}_O\mathbf{Z} + \mathbf{L}\mathbf{D}_O + \mathbf{E}.$$

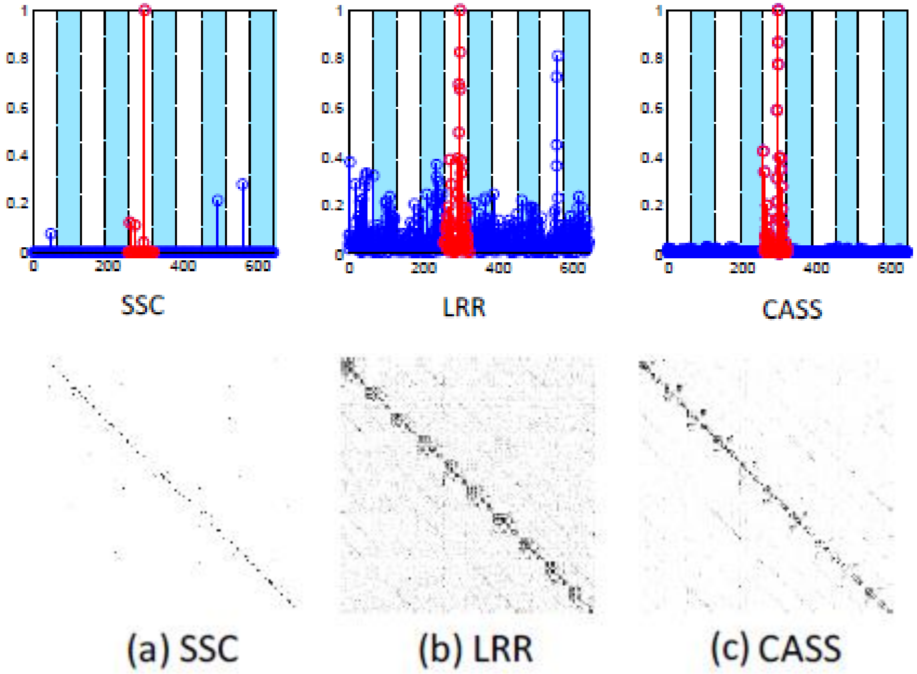
$$\mathbf{D} \quad = \quad \mathbf{DZ} \quad + \quad \mathbf{LD} \quad + \quad \mathbf{E}$$



Liu and Yan. *Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction*, ICCV 2011.

Variants of LRR

- Correlation Adaptive Subspace Clustering



Lu et al. *Correlation Adaptive Subspace Segmentation by Trace Lasso*, ICCV2013.

Variants of LRR

- Correlation Adaptive Subspace Clustering
  - Trace Lasso

$$\|\mathbf{D}\mathrm{diag}(\mathbf{z})\|_*$$

  - Interpolation property

$$\|\mathbf{z}\|_2 \leq \|\mathbf{D}\mathrm{diag}(\mathbf{z})\|_* \leq \|\mathbf{z}\|_1. \quad \text{(columns of } \mathbf{D} \text{ are all normalized to 1)}$$

If data are uncorrelated, i.e., $\mathbf{D}^T\mathbf{D} = \mathbf{I}$, then

$$\|\mathbf{D}\mathrm{diag}(\mathbf{z})\|_* = \|\mathbf{z}\|_1.$$

Adaptive to Data Correlation!

If data are highly correlated, i.e., $\mathbf{D} = \mathbf{u}\mathbf{v}^T$, where $\|\mathbf{u}\|_2 = 1$ and $\mathbf{v}_i = \pm 1$, then

$$\|\mathbf{D}\mathrm{diag}(\mathbf{z})\|_* = \|\mathbf{z}\|_2.$$

Lu et al. *Correlation Adaptive Subspace Segmentation by Trace Lasso*, ICCV2013.
Grave et al. *Trace Lasso: a trace norm regularization for correlated designs*. NIPS2011.

Variants of LRR

- Correlation Adaptive Subspace Clustering

$$\min_{\mathbf{Z}_i} \|\mathbf{D}\mathrm{diag}(\mathbf{Z}_i)\|_*, \quad s.t. \quad \mathbf{D}_i = \mathbf{D}\mathbf{Z}_i, \quad i = 1, 2, \cdots, n.$$

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_n).$$

Lu et al. *Correlation Adaptive Subspace Segmentation by Trace Lasso*, ICCV2013.

Outline

- Representative Models
  - Low-Rank Representation (LRR) Based Subspace Clustering
  - Variants of LRR
- Analysis on LRR
  - Closed-form Solutions in Noiseless Case
  - Exact Recoverability - Deterministic Analysis
  - Exact Recoverability - Probabilistic Analysis
- Applications
- Block-Diagonal Structure in Subspace Clustering

Closed-form Solutions in Noiseless Case

- Noiseless LRR

$$\text{(Original Noiseless LRR)} \quad \min_{\mathbf{Z}} \operatorname{rank}(\mathbf{Z}), \quad s.t. \quad \mathbf{D} = \mathbf{A}\mathbf{Z}. \tag{1}$$

Theorem: Suppose $\mathbf{U}_D\mathbf{\Sigma}_D\mathbf{V}_D^T$ and $\mathbf{U}_A\mathbf{\Sigma}_A\mathbf{V}_A^T$ are the skinny SVD of $\mathbf{D}$ and $\mathbf{A}$, respectively. The complete solutions to problem (1) are given by

$$\mathbf{Z}^* = \boxed{\mathbf{A}^\dagger\mathbf{D}} + \mathbf{S}\mathbf{V}_D^T, \tag{2}$$

where $\mathbf{S}$ is any matrix such that $\mathbf{V}_A^T\mathbf{S} = 0$.

H. Zhang, Z. Lin and C. Zhang, *A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank*, ECML/PKDD 2013.

Closed-form Solutions in Noiseless Case

- Noiseless LRR

$$\text{(Relaxed Noiseless LRR)} \quad \min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D} = \mathbf{AZ}. \tag{3}$$

Theorem: Problem (3) has a unique solution:

$$\mathbf{Z}^* = \mathbf{A}^\dagger \mathbf{D}. \tag{4}$$

Valid for Any Unitary Invariant Norm!

$$\|\mathbf{X}\|_{UI} = \|\mathbf{U}\mathbf{X}\mathbf{V}^T\|_{UI}, \quad \forall\ \mathbf{U}^T\mathbf{U} = \mathbf{I},\ \mathbf{V}^T\mathbf{V} = \mathbf{I}.$$

S. Wei and Z. Lin, *Analysis and improvement of low rank representation for subspace segmentation*, arXiv preprint arXiv:1107.1561.
Liu et al., *Robust Recovery of Subspace Structures by Low-Rank Representation*, IEEE T. PAMI 2013.
H. Zhang, Z. Lin and C. Zhang, *A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank*, ECML/PKDD 2013.
Y.-L. Yu and D. Schuurmans, *Rank/Norm Regularization with Closed-Form Solutions: Application to Subspace Clustering*, UAI2011.

Closed-form Solutions in Noiseless Case

- In Particular

$$\text{(Relaxed Noiseless LRR)} \quad \min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D} = \mathbf{DZ}. \tag{5}$$

Theorem: Let $\mathbf{D} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T$ be the skinny SVD of $\mathbf{D}$, then problem (5) has a unique solution:

$$\mathbf{Z}^* = \mathbf{V}_r \mathbf{V}_r^T. \tag{6}$$

Shape Interaction Matrix

$$\boxed{\min_{\mathbf{D}=\mathbf{DZ}} \|\mathbf{Z}\|_* = \operatorname{rank}(\mathbf{D}).}$$

S. Wei and Z. Lin, *Analysis and improvement of low rank representation for subspace segmentation*, arXiv preprint arXiv:1107.1561.

Deduction of Latent LRR

- Consider unobserved samples

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D}_O = [\mathbf{D}_O, \mathbf{D}_H]\mathbf{Z}. \tag{7}$$

Lemma: Let the skinny SVD of $[\mathbf{D}_O, \mathbf{D}_H]$ be $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and partition $\mathbf{V}$ as $\mathbf{V} = [\mathbf{V}_O; \mathbf{V}_H]$. The solution to (7) is

$$\mathbf{Z}^* = \mathbf{V}\mathbf{V}_O^T.$$

Proof: That $[\mathbf{D}_O, \mathbf{D}_H] = \mathbf{U}\mathbf{\Sigma}[\mathbf{V}_O; \mathbf{V}_H]^T$ implies

$$\mathbf{D}_O = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_O^T, \quad \mathbf{D}_H = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_H^T.$$

So $\mathbf{D}_O = [\mathbf{D}_O, \mathbf{D}_H]\mathbf{Z}$ reduces to:

$$\mathbf{V}_O^T = \mathbf{V}^T\mathbf{Z}.$$

Thus $\mathbf{Z}^* = \mathbf{V}\mathbf{V}_O^T = [\mathbf{V}_O\mathbf{V}_O^T; \mathbf{V}_H\mathbf{V}_O^T]$.

Liu and Yan. *Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction*, ICCV 2011.

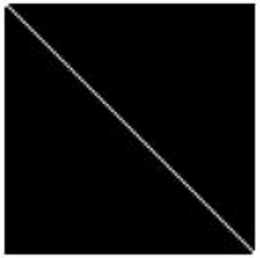
Deduction of Latent LRR

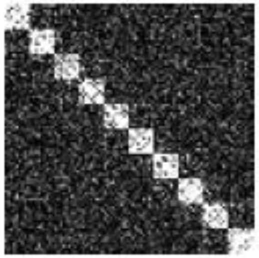
$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D}_O = [\mathbf{D}_O, \mathbf{D}_H]\mathbf{Z}.$$

$$\mathbf{Z}^* = \begin{bmatrix} \mathbf{V}_O\mathbf{V}_O^T \\ \mathbf{V}_H\mathbf{V}_O^T \end{bmatrix}.$$

$$\begin{aligned}
\mathbf{D}_O &= [\mathbf{D}_O, \mathbf{D}_H]\mathbf{Z}^* \\
&= \mathbf{D}_O\mathbf{V}_O\mathbf{V}_O^T + \mathbf{D}_H\mathbf{V}_H\mathbf{V}_O^T \\
&= \mathbf{D}_O\mathbf{V}_O\mathbf{V}_O^T + \mathbf{U}\mathbf{\Sigma}\mathbf{V}_H^T\mathbf{V}_H\mathbf{V}_O^T \\
&= \mathbf{D}_O\underbrace{\boxed{\mathbf{V}_O\mathbf{V}_O^T}}_{\mathbf{Z}} + \underbrace{\boxed{\mathbf{U}\mathbf{\Sigma}\mathbf{V}_H^T\mathbf{V}_H\mathbf{\Sigma}^{-1}\mathbf{U}^T}}_{\mathbf{L}}\mathbf{D}_O.
\end{aligned}$$

low rank!





$$\min_{\mathbf{Z},\mathbf{L}} \operatorname{rank}(\mathbf{Z}) + \operatorname{rank}(\mathbf{L}), \quad s.t.\ \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{L}\mathbf{D}.$$

$$\Rightarrow$$

$$\min_{\mathbf{Z},\mathbf{L}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_*, \quad s.t.\ \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{L}\mathbf{D}.$$

Liu and Yan. *Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction*, ICCV 2011.

Closed-form Solutions in Noiseless Case

- Noiseless Latent LRR

$$\text{(Original Noiseless LatLRR)} \quad \min_{\mathbf{Z},\mathbf{L}} \operatorname{rank}(\mathbf{Z}) + \operatorname{rank}(\mathbf{L}), \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{LD}. \tag{8}$$

$$\Downarrow$$

$$\min_{\mathbf{Z}} \operatorname{rank}(\mathbf{Z}), \quad s.t. \quad \alpha\mathbf{D} = \mathbf{DZ}.$$

$$\min_{\mathbf{L}} \operatorname{rank}(\mathbf{L}), \quad s.t. \quad \beta\mathbf{D} = \mathbf{LD}. \qquad \alpha + \beta = 1.$$

Theorem: The complete solutions to (8) are as follows:

$$\mathbf{Z}^* = \boxed{\mathbf{V}_D\tilde{\mathbf{W}}\mathbf{V}_D^T} + \mathbf{S}_1\tilde{\mathbf{W}}\mathbf{V}_D^T \text{ and } \mathbf{L}^* = \boxed{\mathbf{U}_D\mathbf{\Sigma}_D(\mathbf{I}-\tilde{\mathbf{W}})\mathbf{\Sigma}_D^{-1}\mathbf{U}_D^T} + \mathbf{U}_D\mathbf{\Sigma}_D(\mathbf{I}-\tilde{\mathbf{W}})\mathbf{S}_2,$$

where $\tilde{\mathbf{W}}$ is any $\boxed{\text{idempotent}}$ matrix and $\mathbf{S}_1$ and $\mathbf{S}_2$ are any matrices satisfying:

1. $\mathbf{V}_D^T\mathbf{S}_1 = 0$ and $\mathbf{S}_2\mathbf{U}_D = 0$; and
2. $\operatorname{rank}(\mathbf{S}_1) \leq \operatorname{rank}(\tilde{\mathbf{W}})$ and $\operatorname{rank}(\mathbf{S}_2) \leq \operatorname{rank}(\mathbf{I} - \tilde{\mathbf{W}})$.

$\boxed{\mathbf{A}^2 = \mathbf{A}}$

H. Zhang, Z. Lin and C. Zhang, *A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank*, ECML/PKDD 2013.

Closed-form Solutions in Noiseless Case

- Noiseless Latent LRR

(Relaxed Noiseless LatLRR) $$\min_{\mathbf{Z},\mathbf{L}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_*, \quad s.t. \quad \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{L}\mathbf{D}. \quad (9)$$

Theorem: The complete solutions to (9) are as follows:

$$\mathbf{Z}^* = \mathbf{V}_D \widehat{\mathbf{W}} \mathbf{V}_D^T \text{ and } \mathbf{L}^* = \mathbf{U}_D (\mathbf{I} - \widehat{\mathbf{W}}) \mathbf{U}_D^T,$$

where $\widehat{\mathbf{W}}$ is any block diagonal matrix satisfying:

1. its blocks are compatible with $\mathbf{\Sigma}_D$, i.e., if $[\mathbf{\Sigma}_D]_{ii} \neq [\mathbf{\Sigma}_D]_{jj}$ then $[\widehat{\mathbf{W}}]_{ij} = 0$;
2. both $\widehat{\mathbf{W}}$ and $\mathbf{I} - \widehat{\mathbf{W}}$ are positive semi-definite.

$$\Sigma_X = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix} \Rightarrow \widehat{W} = \begin{bmatrix} \boxed{0.8} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.1 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0.6 & 0.2 & 0 & 0 \\ 0 & 0.3 & 0.2 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boxed{0.3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

Positive Semi-definite

H. Zhang, Z. Lin and C. Zhang, *A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank*, ECML/PKDD 2013.

Relationship between Some Low-Rank Models

(Original Robust LRR) $$\min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda||\mathbf{E}||_\ell, \quad \text{s.t. } \mathbf{D} - \mathbf{E} = (\mathbf{D} - \mathbf{E})\mathbf{Z}.$$

$$\mathbf{A} = \mathbf{D} - \mathbf{E}.$$

$$\min_{\mathbf{Z},\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda||\mathbf{E}||_\ell, \quad \text{s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E}, \ \mathbf{A} = \mathbf{A}\mathbf{Z}.$$

$$\operatorname{rank}(\mathbf{A}) = \min_{\mathbf{A}=\mathbf{A}\mathbf{Z}} \operatorname{rank}(\mathbf{Z}).$$

(Original RPCA) $$\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda||\mathbf{E}||_\ell, \quad \text{s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

$$\min_{\mathbf{Z}} \operatorname{rank}(\mathbf{Z}), \quad \text{s.t. } \mathbf{A} = \mathbf{A}\mathbf{Z}.$$ (Original Noiseless LRR)

Solve First

Use Closed-Form Solution

H. Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.

Relationship between Some Low-Rank Models

$$\text{(Original Robust LRR)} \quad \min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda\|\mathbf{E}\|_\ell, \quad \text{s.t. } \mathbf{D}-\mathbf{E} = (\mathbf{D}-\mathbf{E})\mathbf{Z}.$$

$$\text{(Original Robust LatLRR)} \quad \min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \operatorname{rank}(\mathbf{L}) + \lambda\|\mathbf{E}\|_\ell, \quad \text{s.t. } \mathbf{D}-\mathbf{E} = (\mathbf{D}-\mathbf{E})\mathbf{Z} + \mathbf{L}(\mathbf{D}-\mathbf{E}).$$

$$\text{(Relaxed Robust LRR)} \quad \min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_\ell, \quad \text{s.t. } \mathbf{D}-\mathbf{E} = (\mathbf{D}-\mathbf{E})\mathbf{Z}.$$

$$\text{(Relaxed Robust LatLRR)} \quad \min_{\mathbf{Z},\mathbf{L},\mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda\|\mathbf{E}\|_\ell, \quad \text{s.t. } \mathbf{D}-\mathbf{E} = (\mathbf{D}-\mathbf{E})\mathbf{Z} + \mathbf{L}(\mathbf{D}-\mathbf{E}).$$

$$\Downarrow$$

$$\text{(Original RPCA)} \quad \min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda\|\mathbf{E}\|_\ell, \quad \text{s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

$+$ Closed-Form Solutions

H. Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.

Relationship between Some Low-Rank Models

Original Robust LRR

Relaxed Robust LRR

RPCA

Original Robust LatLRR

Relaxed Robust LatLRR

$$\text{(Original RPCA)} \quad \min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda ||\mathbf{E}||_{\ell}, \quad \text{s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

H. Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.

Implications

- We could obtain a *globally optimal* solution to other low rank models.
- We could have *much faster* algorithms for other low rank models.

$$\text{(Original RPCA)} \quad \min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda ||\mathbf{E}||_1, \quad \text{s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

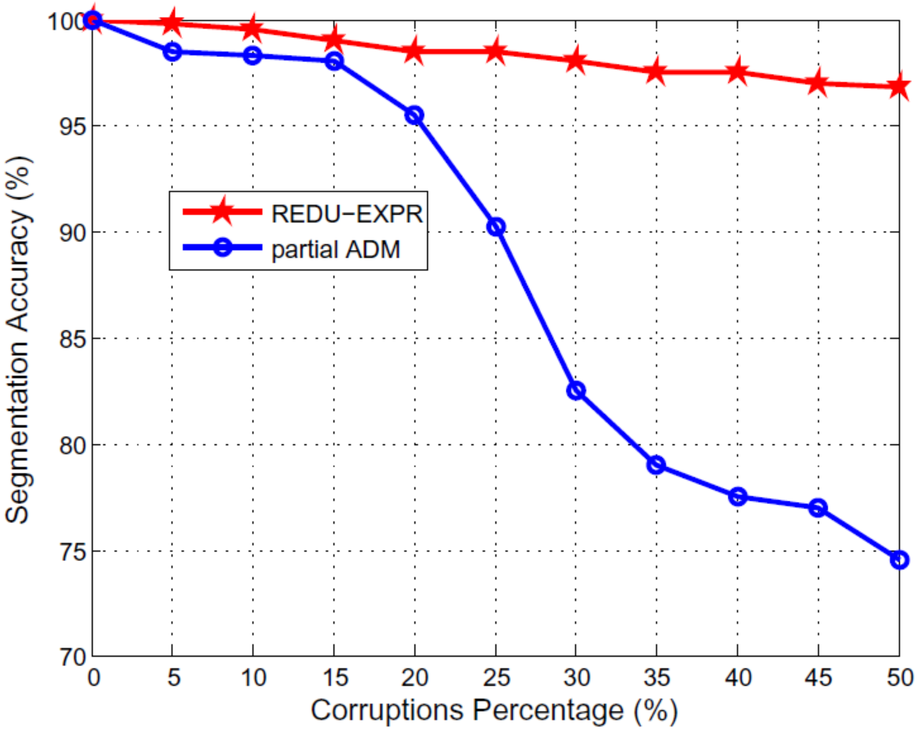
$$\Downarrow$$

$$\text{(Relaxed RPCA)} \quad \min_{\mathbf{A},\mathbf{E}} ||\mathbf{A}||_* + \lambda ||\mathbf{E}||_1, \quad \text{s.t. } \mathbf{D} = \mathbf{A} + \mathbf{E}.$$

H. Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.
E. Candes et al. *Robust Principal Component Analysis?* J. ACM, 2011.

Implications

- Comparison of optimality



REDU-EXPR = Reduce to RPCA and then use closed-form solution

partial ADM = Solve the problem directly

Comparison of accuracies of solutions to relaxed R-LRR computed by REDU-EXPR and partial ADM, where the parameter is adopted as $1/\sqrt{\log n}$ and n is the input size. The program is run by 10 times and the average accuracies are reported.

H. Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.

Implications

- Comparison of speed

Model	Method	Accuracy	CPU Time (h)
LRR	ADM	-	>10
R-LRR	ADM	-	did not converge
R-LRR	partial ADM	-	>10
R-LRR	REDU-EXPR	61.6365%	0.4603

Table 2: Unsupervised face image clustering results on the Extended YaleB database. REDU-EXPR means reducing to RPCA first and then express the solution as that of RPCA.

H. Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.

Exact Recoverability - Deterministic Analysis

Theorem: Let $\lambda = 3/(7\|\mathbf{D}\|\sqrt{\gamma n})$. Then there exists $\gamma^* > 0$ such that when $\gamma \leq \gamma^*$, LRR can exactly recover the **row space** and the **column support** of $(\mathbf{Z}_0, \mathbf{E}_0)$:

$$\mathbf{U}^*(\mathbf{U}^*)^T = \mathbf{V}_0\mathbf{V}_0^T, \quad \mathcal{I}^* = \mathcal{I}_0,$$

where $\gamma = |\mathcal{I}_0|/n$ is the fraction of outliers, $\mathbf{U}^*$ is the column space of $\mathbf{Z}^*$, $\mathbf{V}_0^T$ is the row space of $\mathbf{Z}_0$, and $\mathcal{I}^*$ and $\mathcal{I}_0$ is the column supports of $\mathbf{E}^*$ and $\mathbf{E}_0$, respectively.

The magnitude of outliers does not matter!

Liu et al., *Exact Subspace Segmentation and Outlier Detection by Low-Rank Representation*, AISTATS 2012.
Liu et al., *Robust Recovery of Subspace Structures by Low-Rank Representation*, IEEE T. PAMI 2013.
Liu et al., *A Deterministic Analysis for LRR*, IEEE T. PAMI, 2016.

Exact Recoverability – Probabilistic Analysis

(Original Robust LRR) $\min_{\mathbf{Z},\mathbf{E}} \operatorname{rank}(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{2,0}, \quad \text{s.t.} \quad \mathbf{D} - \mathbf{E} = (\mathbf{D} - \mathbf{E})\mathbf{Z}.$

$\Downarrow$

(Original Outlier Pursuit) $\min_{\mathbf{A},\mathbf{E}} \operatorname{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_{2,0}, \quad \text{s.t.} \quad \mathbf{D} = \mathbf{A} + \mathbf{E}.$

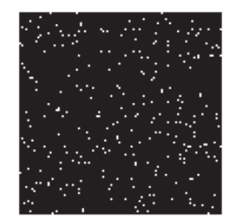
$\Downarrow$ + Closed-Form Solutions

(Relaxed Outlier Pursuit) $\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \quad \text{s.t.} \quad \mathbf{D} = \mathbf{A} + \mathbf{E}.$

Column Corruption

(Relaxed RPCA) $\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1, \quad \text{s.t.} \quad \mathbf{D} = \mathbf{A} + \mathbf{E}.$

Entry Corruption



Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.

Zhang et al. *Exact Recoverability of Robust PCA via Outlier Pursuit with Tight Recovery Bounds*, AAAI2015.

Zhang et al. *Completing Low-Rank Matrices with Corrupted Samples from Few Coefficients in General Basis*, IEEE T. Information Theory, 2016.

Y. Chen et al. *Robust matrix completion and corrupted columns*, ICML 2011.

E. Candes et al. *Robust Principal Component Analysis?* J. ACM, 2011.

Exact Recoverability – Probabilistic Analysis

Theorem: Suppose $m = \Theta(n)$, $\text{Range}(\mathbf{L}_0) = \text{Range}(\mathcal{P}_{\mathcal{I}_0^\perp}\mathbf{L}_0)$, and $[\mathbf{S}_0]_{:,j} \notin \text{Range}(\mathbf{L}_0)$ for $\forall j \in \mathcal{I}_0$. Then any solution $(\mathbf{L}_0 + \mathbf{H}, \mathbf{S}_0 - \mathbf{H})$ to Outlier Pursuit with $\boxed{\lambda = 1/\sqrt{\log n}}$ exactly recovers the column space of $\mathbf{L}_0$ and the comumn support of $\mathbf{S}_0$ with a probability at least $1 - cn^{-10}$, if the column support $\mathcal{I}_0$ of $\mathbf{S}_0$ is uniformly distributed among all sets of cardinality s and

$$\text{rank}(\mathbf{L}_0) \le \rho_r \frac{\min(m,n)}{\mu \log n} \quad \text{and} \quad s \le \rho_s n,$$

where c, ρ_r and ρ_s are constants, and other conditions to avoid column sparsity and clustering of outier columns.

Also good in deterministic analysis

H. Zhang et al., *Exact Recoverability of Robust PCA via Outlier Pursuit with Tight Recovery Bounds*, AAAI2015.

H. Zhang et al., *Completing Low-Rank Matrices with Corrupted Samples from Few Coefficients in General Basis*, IEEE T. Information Theory, 2016.

Liu et al., *A Deterministic Analysis for LRR*, IEEE T. PAMI, 2016.

Outline

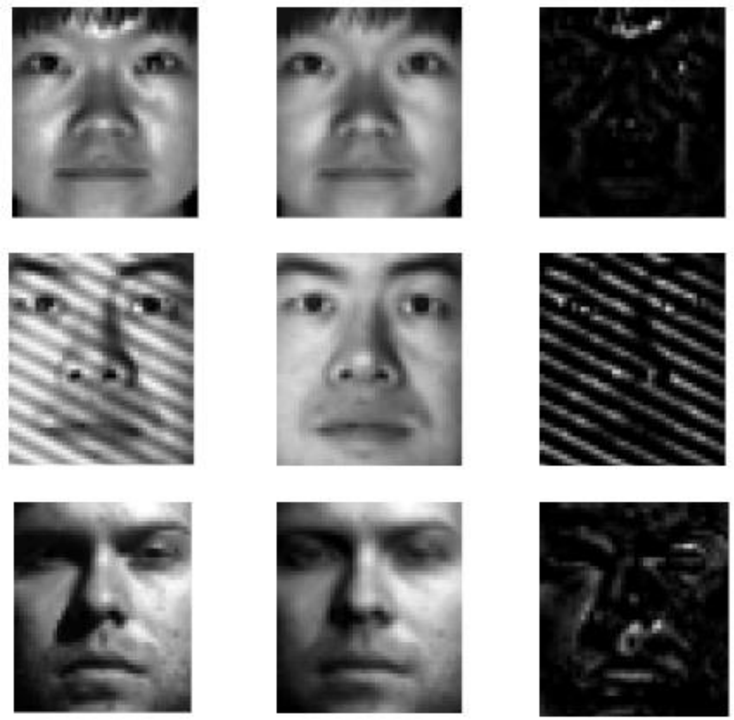
- Representative Models
  - Low-Rank Representation (LRR) Based Subspace Clustering
  - Variants of LRR
- Analysis on LRR
  - Closed-form Solutions in Noiseless Case
  - Exact Recoverability - Deterministic Analysis
  - Exact Recoverability - Probabilistic Analysis
- Applications
- Block-Diagonal Structure in Subspace Clustering

Facial Image Denoising

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{E}.$$

Images in Columns

$\mathbf{D} = \mathbf{DZ} + \mathbf{E}$

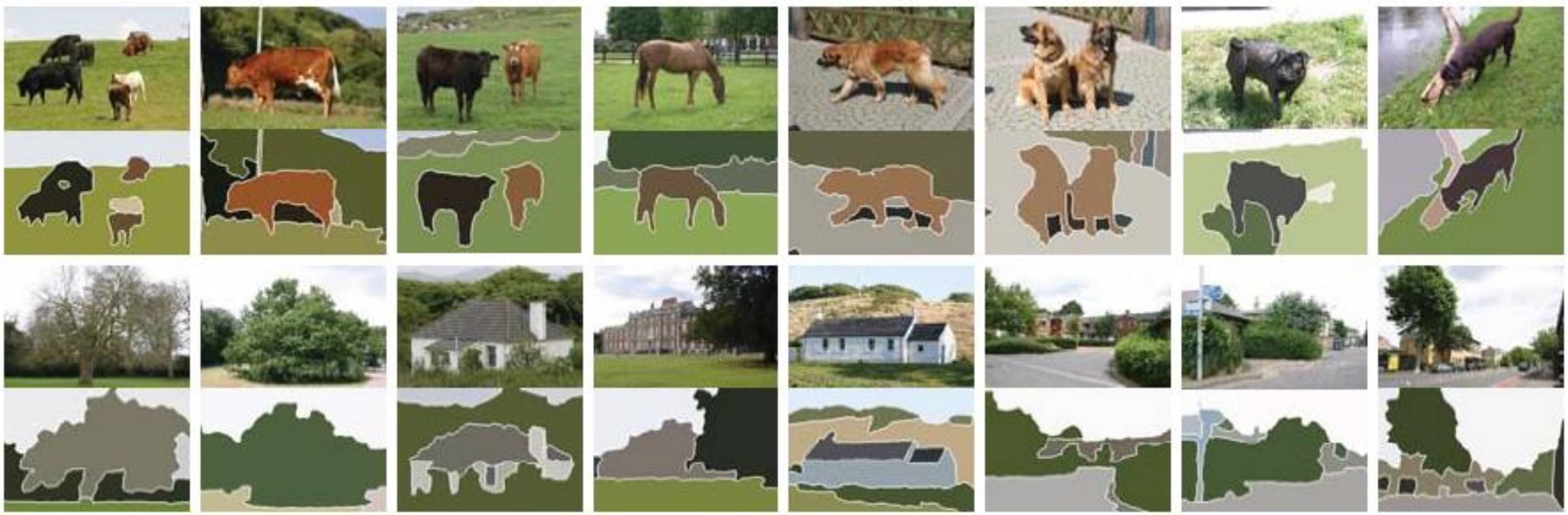


Liu et al. *Robust Subspace Segmentation by Low-Rank Representation*, ICML 2010.

Image Segmentation

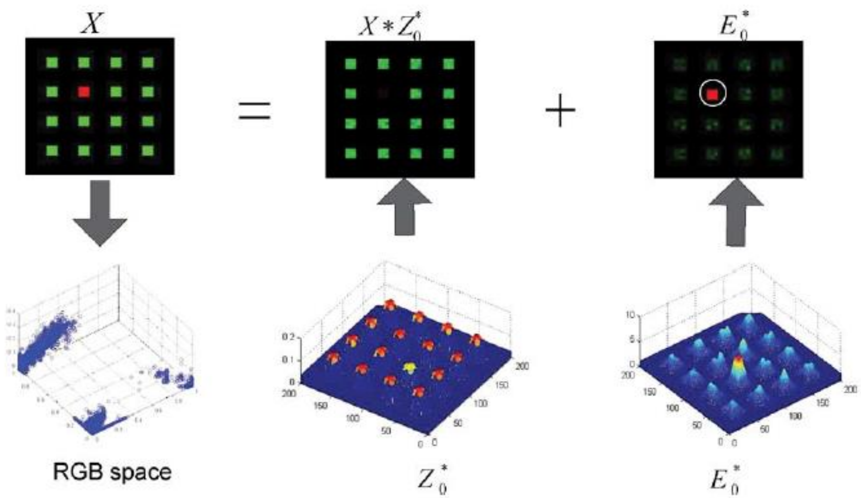
$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{E}.$$

Feature of Superpixel



Cheng et al. *Multi-task Low-rank Affinity Pursuit for Image Segmentation*, ICCV 2011.

Saliency Detection



$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1},$$
$$s.t. \quad \mathbf{D} = \mathbf{D}\mathbf{Z} + \mathbf{E}.$$

Features of Image Patches

Lang et al. *Saliency Detection by Multitask Sparsity Pursuit*. IEEE TIP 2012.

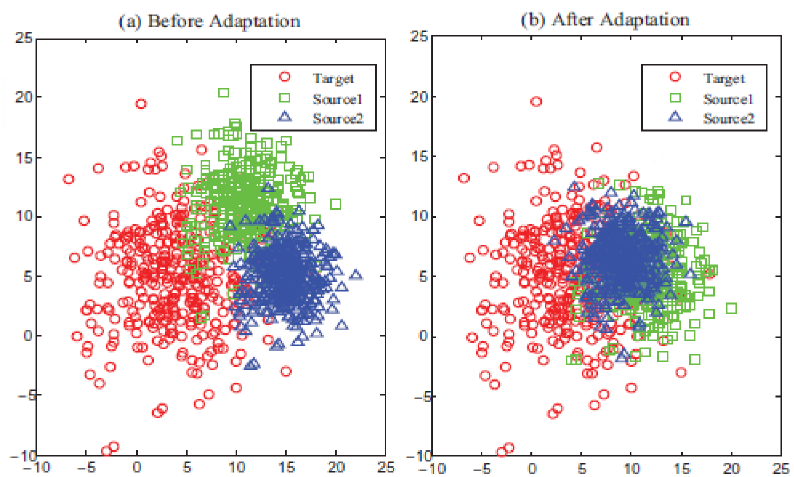
Visual Domain Adaptation

$$\min_{\mathbf{W},\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{WS} = \mathbf{TZ}+\mathbf{E},\ \ \mathbf{WW}^T=\mathbf{I}.$$

Transformation Matrix

Source Feature

Target Feature



Bookcase images of different domains

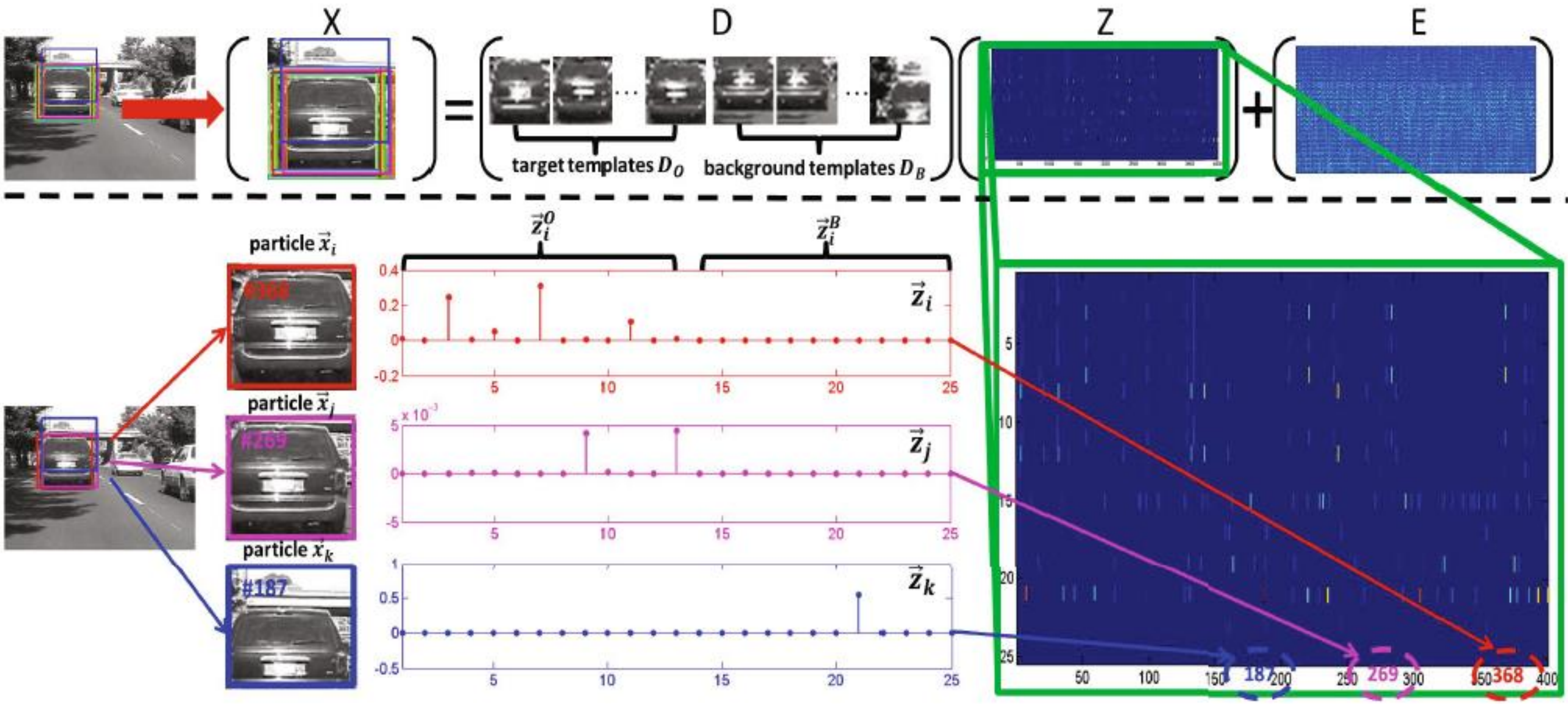
Jhuo et al., *Robust Visual Domain Adaptation with Low-Rank Reconstruction*, CVPR2012.

Robust Visual Tracking

$$\min_{\mathbf{Z},\mathbf{E}} \lambda_1 \|\mathbf{Z}\|_* + \lambda_2 \|\mathbf{Z}\|_1 + \lambda_3 \|\mathbf{E}\|_1, \quad s.t. \quad \mathbf{X} = \mathbf{D}_t \mathbf{Z} + \mathbf{E}.$$

Observation of Particle

Templates of Obj. and Bkgd.



Zhang et al., *Low-Rank Sparse Learning for Robust Visual Tracking*, ECCV2012.

Extract Feature for 3D Face Recognition

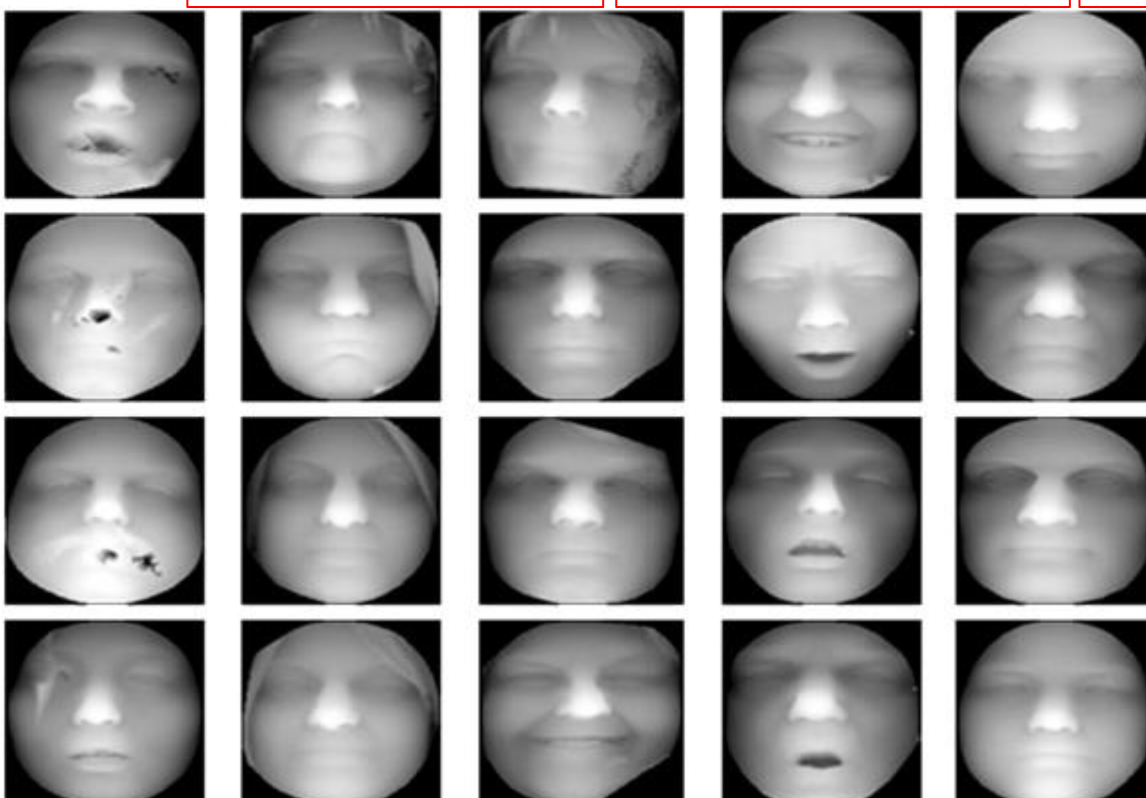
Bounding Sphere Representation

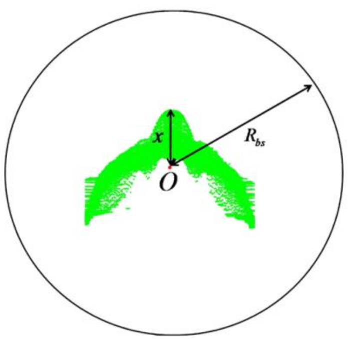
$$\min_{\mathbf{R},\mathbf{X},\mathbf{E}} \frac{1}{2}\|\mathbf{R}\|_F^2 + \lambda_1\|\mathbf{X}\|_* + \lambda_2\|\mathbf{E}\|_1, \quad s.t. \quad \mathbf{Y} = \mathbf{R}\mathbf{X},\ \mathbf{S} = \mathbf{X} + \mathbf{E}.$$

Response Matrix

Regression Matrix

Explanatory Matrix





Ming and Ruan, *Robust sparse bounding sphere for 3D face recognition*, IVC, 2012.

Computed Tomography

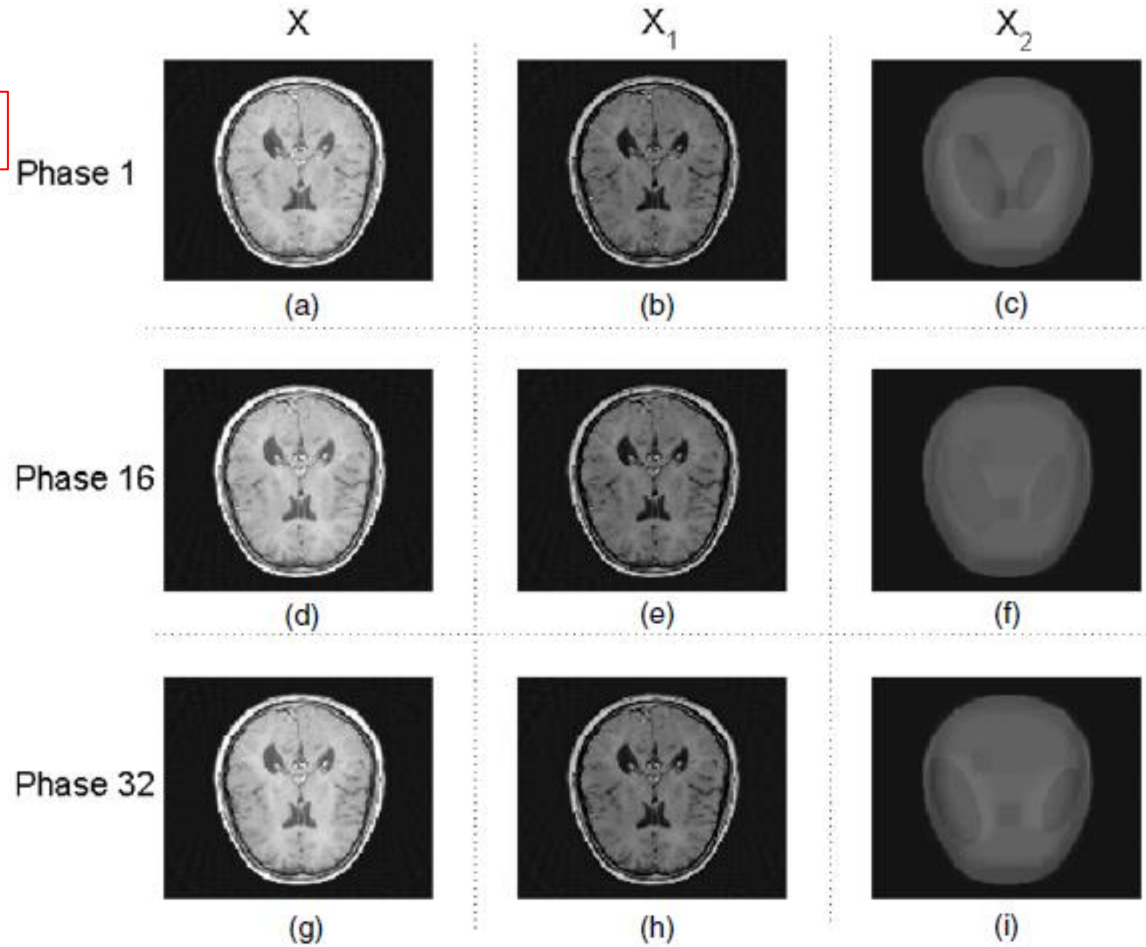
Integration Matrices

Measurements

$$\min_{\mathbf{X}_1,\mathbf{X}_2} \|\mathbf{A}(\mathbf{X}_1+\mathbf{X}_2)-\mathbf{Y}\|_F^2$$

$$+\lambda_1\|\mathbf{X}_1\|_* + \lambda_2\|\mathbf{W}\mathbf{X}_2\|_1.$$

Framelets



Gao et al., *Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)*, Inverse Problems, 2011.

Gao et al., *Robust principal component analysis-based four-dimensional computed tomography*, Phys. Med. Biol., 2011.

More Applications

- **Scene Parsing**
  - Xuelong Li, Lichao Mou, and Xiaoqiang Lu. *Scene parsing from an map perspective*. IEEE T. Cybernetics, 2015
- **Image Shrinkage**
  - Qi Wang, and Xuelong Li. *Shrink image by feature matrix decomposition*. Neurocomputing, 2014.
- **Optic Disc Detection**
  - Huazhu Fu, Dong Xu, Stephen Lin, Damon Wing Kee Wong, Jiang Liu. *Automatic Optic Disc Detection in OCT Slices via Low-Rank Reconstruction*. IEEE T. Bio-medical Engineering, 2014.
- **Image Classification**
  - Geoff Bull and Junbin Gao. *Transposed Low Rank Representation for Image Classification*. In Digital Image Computing Techniques and Applications (DICTA), 2012.
- **Image Representation**
  - Xue Li, Hongxun Yao, Xiaoshuai Sun, and Yanhao Zhang. *On Dense Sampling Size*. ICIP 2013.
- **Image Annotation**
  - Teng Li, Bin Cheng, Xinyu Wu, Jun Wu. *Low-Rank Affinity Based Local-Driven Multi-label Propagation*. Mathematical Problems in Engineering, 2013.
- **Video Summarization**
  - Zhengzheng Tu, Dengdi Sun, and Bin Luo. *Video Summarization by Robust Low-Rank Subspace Segmentation*. In Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013.
- **Semantic Concept Detection**
  - Meng Jian, Cheolkon Jung, and Yaoguo Zheng, *Discriminative Structure Learning for Semantic Concept Detection with Graph Embedding*, IEEE T. Multimedia, 2014.
- **Dimensionality Reduction**
  - Chun-Guang Li, Xianbiao Qi, and Jun Guo. *Dimensionality reduction by low-rank embedding*. Intelligent Science and Intelligent Data Engineering, 2013.
- **Gene Clustering**
  - Yan Cui, Chun-Hou Zheng, and Jian Yang. *Identifying Subspace Gene Clusters from Microarray Data Using Low-Rank Representation*. PloS One, 2013.
- **Music Analysis**
  - Yannis Panagakis and Constantine Kotropoulos. *Automatic Music Mood Classification via Low-Rank Representation*. European Signal Processing Conference, 2011.
  - Yannis Panagakis and Constantine Kotropoulos. *Automatic Music Tagging by Low-rank Representation*. In IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP), 2012.

References

- Liu et al. *Robust Subspace Segmentation by Low-Rank Representation*, ICML 2010.
- Liu et al. *Robust Recovery of Subspace Structures by Low-Rank Representation*, TPAMI 2013.
- Liu et al. *Fixed-Rank Representation for Unsupervised Visual Learning*, CVPR 2012.
- Liu et al. *Exact Subspace Segmentation and Outlier Detection by Low-Rank Representation*, AISTATS 2012.
- Liu et al., *A Deterministic Analysis for LRR*, IEEE T. PAMI, 2016.
- Liu and Yan. *Latent Low-Rank Representation for Subspace Segmentation and Feature Extraction*, ICCV 2011.
- Lu et al. *Correlation Adaptive Subspace Segmentation by Trace Lasso*, ICCV2013.
- Wei and Lin, *Analysis and Improvement of Low Rank Representation for Subspace Segmentation*, arXiv preprint arXiv:1107.1561.
- Zhang et al. *A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank*, ECML/PKDD 2013.
- Zhang et al. *Exact Recoverability of Robust PCA via Outlier Pursuit with Tight Recovery Bounds*, AAAI2015
- Zhang et al. *Relation among Some Low Rank Subspace Recovery Models*, Neural Computation, 2015.
- Zhang et al. *Completing Low-Rank Matrices with Corrupted Samples from Few Coefficients in General Basis*, IEEE T. Information Theory, 2016.

Outline

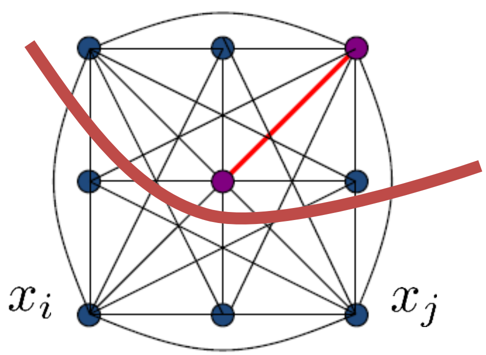
- Representative Models
  - Low-Rank Representation (LRR) Based Subspace Clustering
  - Variants of LRR
- Analysis on LRR
  - Closed-form Solutions in Noiseless Case
  - Exact Recoverability - Deterministic Analysis
  - Exact Recoverability - Probabilistic Analysis
- Applications
- Block-Diagonal Structure in Subspace Clustering

Outline

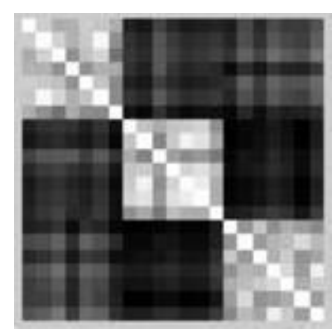
- What is the block diagonal structure?
- What clustering algorithms offer the block diagonal structure?
- How to pursue the block diagonal structure?
- Is the exactly block diagonal structure necessary?
- Conclusions

Spectral Clustering based Methods

- Spectral clustering for subspace clustering (SC)
  - Graph construction
    - Construct a data affinity matrix *W* encoding data points pairwise similarity.
    - Many SC methods focus on this step.
  - Normalize cut over the graph
    - Partition data points into multiple clusters (subspaces).



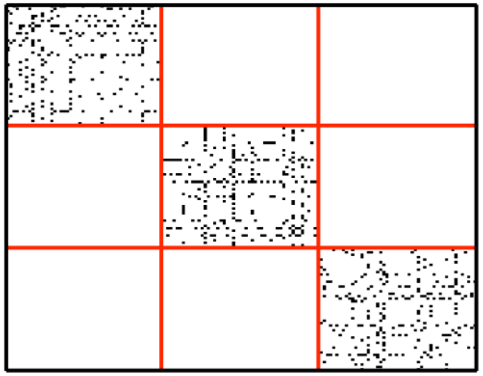
$\mathcal{G} = \langle X, W \rangle$



$W = [w_{ij}]_{n \times n} = [\mathrm{sim}(x_i, x_j)]_{n \times n}$

Spectral Clustering based Methods

- Spectral clustering for subspace clustering
  - Ideally, W is a block-diagonal one.
  - The similarity between data points from different subspaces are all zero.
  - # blocks = # subspaces



After proper arrangement of the data points.

$\boldsymbol{W}_{ij}$ **= 0,** if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are from different subspaces

Sparse Subspace Clustering

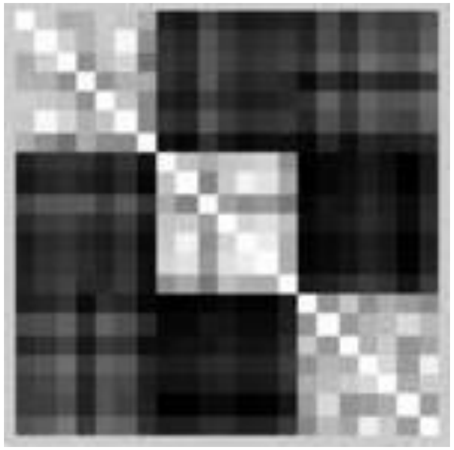
- Sparse subspace clustering (SSC)

$$\min_Z \|Z\|_0, \text{ s.t. } X = XZ, \text{ diag}(Z) = 0.$$

$$\min_Z \|Z\|_1, \text{ s.t. } X = XZ, \text{ diag}(Z) = 0.$$ (convex relaxation)

- Affinity matrix:

$$W = \frac{1}{2}(|Z| + |Z^\top|)$$

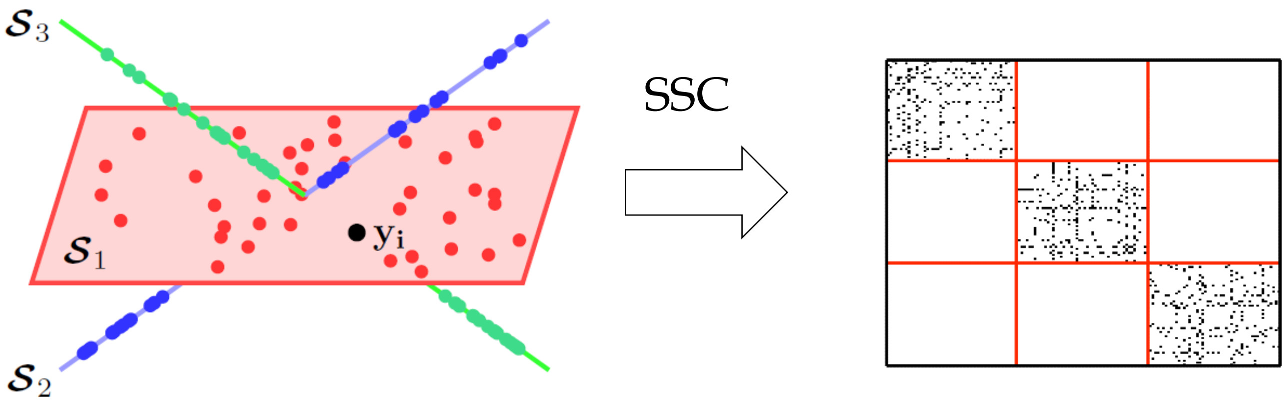


Elhamifar, et al. *Sparse Subspace Clustering*. CVPR 2009
Cheng, et al. *Learning with l1-graph for Image Analysis*. TIP 2010
Elhamifar, et al. *Sparse Subspace Clustering: Algorithm, Theory, and Applications*. TPAMI 2013

Sparse Subspace Clustering

- Independent subspace assumption:
  k subspaces are independent if $\dim(\oplus_{i=1}^{k} \mathcal{S}_i) = \sum_{i=1}^{k} \dim(\mathcal{S}_i)$

- The solution Z to the SSC is block diagonal when the subspaces are independent.



Independent subspaces

Block diagonal affinity matrix

Elhamifar, et al. *Sparse Subspace Clustering*. CVPR 2009
Elhamifar, et al. *Sparse Subspace Clustering: Algorithm, Theory, and Applications*. TPAMI 2013

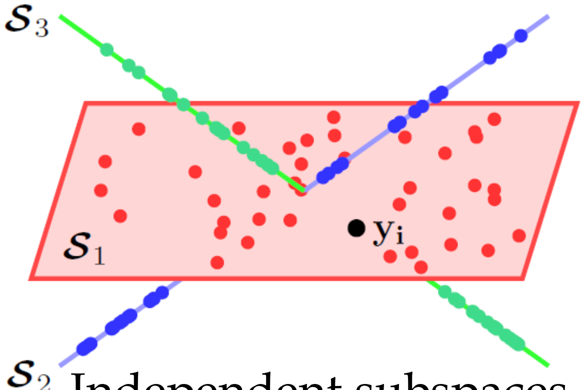
Low-rank Subspace Clustering

- Low-rank representation (LRR)

$$\min_{Z} \ \mathrm{rank}(Z), \ \text{s.t.} \ \ X = XZ.$$

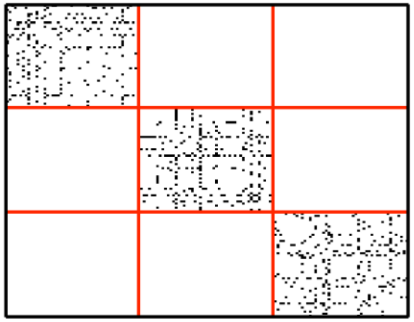
$$\min_{Z} \ \|Z\|_*, \ \text{s.t.} \ \ X = XZ. \qquad \text{(convex relaxation)}$$

- The solution Z to the LRR is block diagonal when the subspaces are independent.



Independent subspaces

LRR



Block diagonal affinity matrix

Liu, et al. *Robust Subspace Segmentation by Low-Rank Representation*. ICML 2010

Liu, et al. *Robust Recovery of Subspace Structures by Low-Rank Representation*. TPAMI 2013

Other SC Methods

- Multiple-subspace representation (MSR)

$$\min_{Z} \|Z\|_* + \lambda \|Z\|_1, \quad \text{s.t.} \quad X = XZ, \ \mathrm{diag}(Z) = 0.$$

- Subspace segmentation via quad. prog. (SSQP)

$$\min_{Z} \|Z^\top Z\|_1, \quad \text{s.t.} \ X = XZ, \quad Z \geq 0, \ \mathrm{diag}(Z) = 0.$$

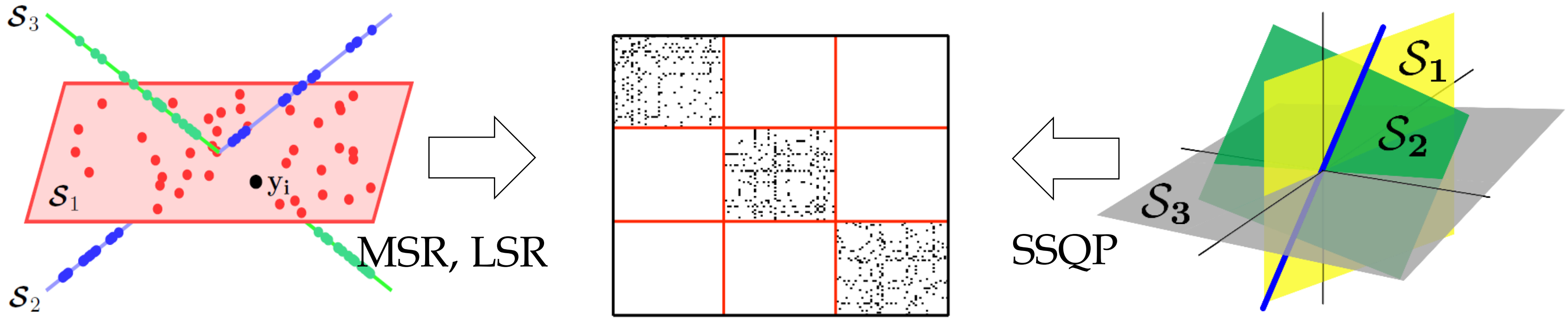
- Least squares regression (LSR)

$$\min_{Z} \|Z\|_F, \quad \text{s.t.} \quad X = XZ.$$

Luo, et al. *Multi-Subspace Representation and Discovery*. ECML PKDD 2011
Wang, et al. *Efficient Subspace Segmentation via Quadratic Programming*, AAAI 2011
Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012

BD Properties of Other SC methods

- If the subspaces are independent, the solutions to MSR and LSR are block diagonal.
- If the subspaces are orthogonal, the solution to SSQP is block diagonal.



Luo, et al. *Multi-Subspace Representation and Discovery*. ECML PKDD 2011
Wang, et al. *Efficient Subspace Segmentation via Quadratic Programming*, AAAI 2011
Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012

Unifying Previous Methods

- Previous methods can be written as:

$$\min f(Z) \quad \text{s.t. } Z \in \Omega$$

	$f(Z)$	Ω
SSC	$\|Z\|_0$ or $\|Z\|_1$	$\{Z \mid X = XZ, \ \operatorname{diag}(Z) = 0\}$
LRR	$\|Z\|_*$	$\{Z \mid X = XZ\}$
MSR	$\|Z\|_1 + \lambda\|Z\|_*$	$\{Z \mid X = XZ, \operatorname{diag}(Z) = 0\}$
SSQP	$\|Z^\top Z\|_1$	$\{Z \mid X = XZ, Z \geq 0, \operatorname{diag}(Z) = 0\}$
LSR	$\|Z\|_F$	$\{Z \mid X = XZ\}$
Structured SSC	$\|Z\|_{1,2}$	$\{Z \mid X = XZ, \operatorname{diag}(Z) = 0\}$

- All the above methods give block diagonal solution for either independent or orthogonal subspaces.
- However, they are proved case by case.

What f Gives Block Diagonality?

- Consider the following general SC formulation:

$$\min f(Z) \quad \text{s.t. } Z \in \Omega$$

- What kind of objective functions f induce the block diagonal solution under certain assumption?
  - Orthogonal subspaces
  - Independent subspaces

Enforced Block Diagonal Cond.

Enforced Block Diagonal (EBD) Conditions. Assume the matrix function f is defined on $\Omega(\neq \varnothing)$. For any $Z = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \Omega$, $Z \neq 0$, where A and D are square matrices, B and C are of compatible dimension, $A, D \in \Omega$. Let $Z^D = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \in \Omega$. The EBD conditions are:

(1) $f(Z) = f(P^\top Z P)$, $\forall$ permutation matrix $P \in \{P | P^\top Z P \in \Omega\}$.

(2) $f(Z) \geq f(Z^D)$, where the equality holds iff $B = C = 0$.

(3) $f(Z^D) = f(A) + f(D)$.

Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012

Enforced Block Diagonal Cond.

- EBD conditions
  1. $f(Z) = f(P^T Z P)$
  2. $f(Z) \geq f(Z^D)$
  3. $f(Z^D) = f(A) + f(D)$

$$Z^D = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}$$

List of functions f satisfying EBD conditions.

	$f(Z)$	Ω
SSC	$\|Z\|_0$ or $\|Z\|_1$	$\{Z \mid X = XZ,\ \mathrm{diag}(Z) = 0\}$
MSR	$\|Z\|_1 + \lambda\|Z\|_*$	$\{Z \mid X = XZ, \mathrm{diag}(Z) = 0\}$
SSQP	$\|Z^\top Z\|_1$	$\{Z \mid X = XZ, Z \geq 0, \mathrm{diag}(Z) = 0\}$
LSR	$\|Z\|_F$	$\{Z \mid X = XZ\}$
Structured SSC	$\|Z\|_{1,2}$	$\{Z \mid X = XZ, \mathrm{diag}(Z) = 0\}$
Other choices	$(\sum_{i,j=1}^n \lambda_{ij} \lvert Z_{ij}\rvert^{p_{ij}})^s$ $\lambda_{ij} > 0,\ p_{ij} > 0,\ s > 0$	$\{Z \mid X = XZ, \mathrm{diag}(Z) = 0\}$

Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012

EBD guarantees Block Diagonal: Independent Subspaces

f satisfies EBD cond.
+ → block diagonal Z
subspaces are independent

Theorem Assume data sampling is sufficient, and subspaces are independent. Consider:

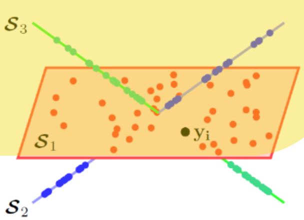
$$\min f(Z) \;\; \text{s.t. } Z \in \Omega = \{Z | X = XZ\}.$$

If f satisfies the EBD conditions (1)(2), then Z^* is block diagonal

$$Z^* = \text{blkdiag}(Z_1^*, Z_2^*, \ldots, Z_k^*)$$

with $Z_i^* \in \mathbb{R}^{n_i \times n_i}$ corresponding to X_i, for each i. Furthermore, if f satisfies the EBD conditions (1)(2)(3), for each i, Z_i^* is optimal solution to:

$$\min f(Y) \;\; \text{s.t. } X_i = X_i Y$$



Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012

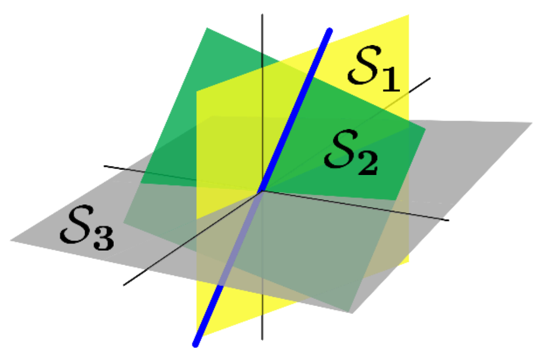
EBD guarantees Block Diagonal: Orthogonal Subspaces

Theorem If the subspaces are orthogonal, and f satisfies EBD conditions (1)(2), the optimal solution(s) to the following problem:

$$\min ||X - XZ||_{2,p} + \lambda f(Z)$$

must be block diagonal, where $\|X\|_{2,p} = (\sum_j (\sum_{i=1}^n X_{ij}^2)^{\frac{p}{2}})^{\frac{1}{p}}$, $p > 0$, and $\lambda > 0$.

- The orthogonal assumption is stronger than the independent assumption. Therefore, easier to get block diagonal solutions.



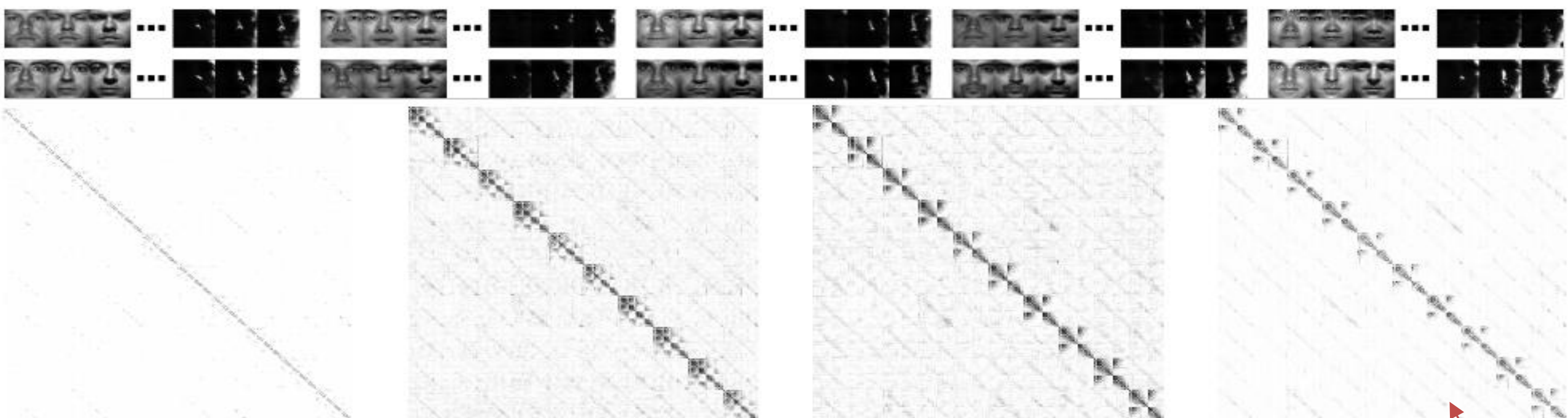
Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012

Trace Lasso based SC

- Correlation Adaptive Subspace Segmentation (CASS).

$$\min_{Z} \sum_{i} \|X \operatorname{diag}(Z_i)\|_*, \quad \text{s.t.} \quad X = XZ.$$

- Trace Lasso: adaptive balance between L_1- and L_2-norm.
- CASS satisfies the EBD conditions*. The solution is block diagonal when the subspaces are independent.



(a) SSC (b) LRR (c) LSR (d) CASS

Clearer BD structure

* More strictly, enforced block sparse (EBS) conditions.

Lu, et al. *Correlation Adaptive Subspace Segmentation by Trace Lasso*. ICCV 2013.

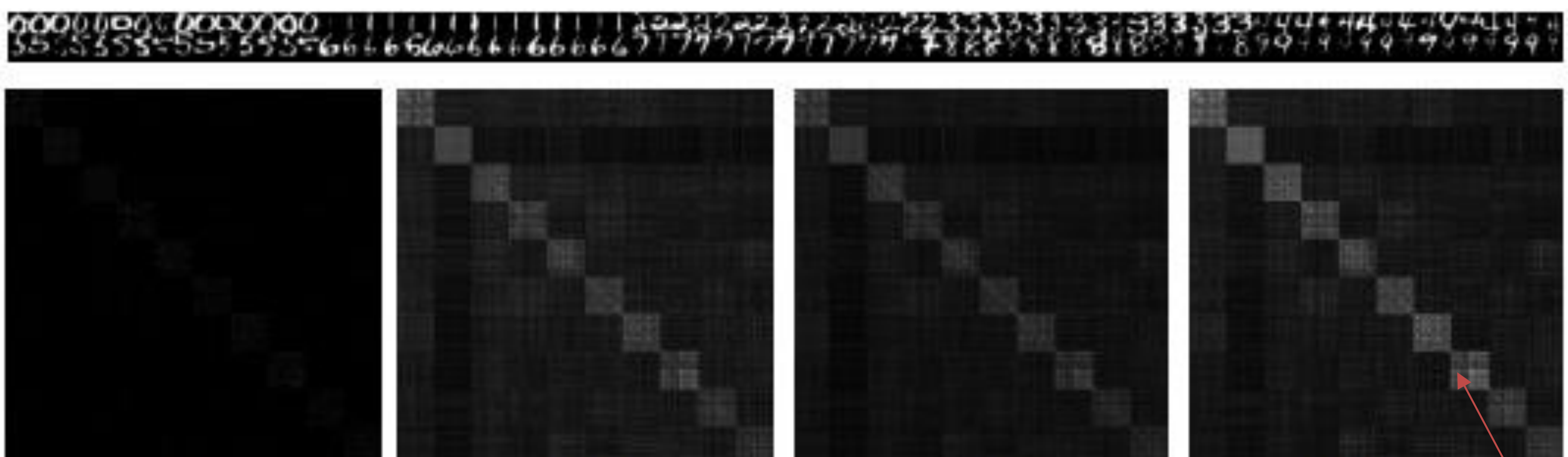
SMooth Representation

- SMooth Representation (SMR):

$$\min_{Z} \mathrm{Tr}(ZLZ^{\top}) \quad \text{s.t.} \quad X = XZ.$$

Here, L is Laplacian matrix of X.

- SMR aims to enhance grouping effect: $x_i \to x_j$, then $z_i \to z_j$.
- SMR satisfies the EBD conditions*.



(a) SSC (b) LRR (c) LSR (d) SMR

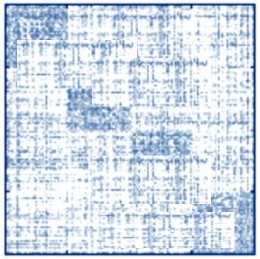
Enhanced intra-block density

* More strictly, enforced grouping effect (EGE) conditions.

Hu et al. *Smooth Representation Clustering*, CVPR2014.

How to Pursue Block-diagonality?

- Self representation: $x_i = Xz_i$
- Priors on $Z = [z_1, \ldots, z_n]$
  - Sparse: $\|Z\|_1 \leq s$, Low-rank: $\|Z\|_* \leq r$
  - Enforced block-diagonal condition
  - **Noiseless & independent** subspace: block-



- NOT block-diagonal in practice $W = (|Z| + |Z^\top|)/2$
  - Noises, non-independent subspaces, too close subspaces

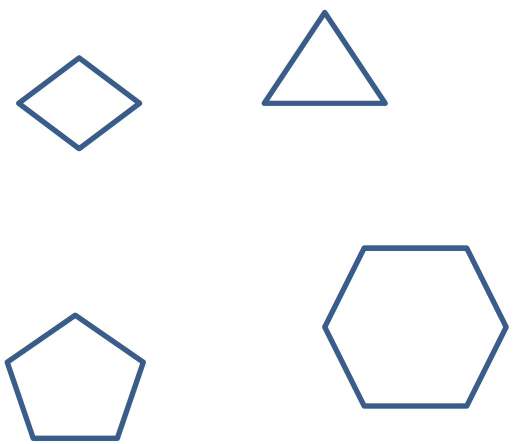
Elhamifar, et al. *Sparse Subspace Clustering*. CVPR 2009
Lu, et al. *Robust and Efficient Subspace Segmentation via Least Squares Regression*. ECCV 2012
Liu, et al. *Robust Recovery of Subspace Structures by Low-Rank Representation*. TPAMI 2013
Elhamifar, et al. *Sparse Subspace Clustering: Algorithm, Theory, and Applications*. TPAMI 2013

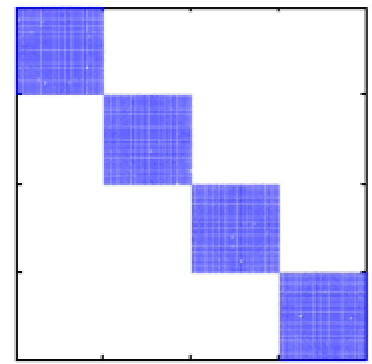
Questions

- Exactly block-diagonal structure prior?
- Efficient optimization?
- Performance guarantee?

Block-diagonal Prior

- Recall a spectral graph theory





Theorem Let W be an affinity matrix. Then the multiplicity k of the eigenvalue 0 of the corresponding Laplacian L_W equals the number of connected components in W.

Here, $L_W(j, j') = -W(j, j'),$ if $j \neq j'$; $\sum_{\ell \neq j} W(j, \ell)$ otherwise.

$$\boxed{W \text{ is } k \text{ block diagonal } \Leftrightarrow W \in \left\{ W \,\middle|\, \operatorname{rank}(L_W) = n - k, W \in \mathbb{R}^{n \times n} \right\}}$$

Explicit block-diagonal prior

- The Block-diagonal prior

$$Z \in \mathcal{K} = \left\{ Z \,\middle|\, \mathrm{rank}(L_W) = n - k, W = (|Z| + |Z^\top|)/2, Z \in \mathbb{R}^{n \times n} \right\}$$

- "Block" is defined as the connected component: no ambiguities
- The first exactly block-diagonal structure prior

Feng, et al. *Robust Subspace Segmentation with Block-Diagonal Prior*. CVPR 2014

Applications of BD Matrix Pursuit

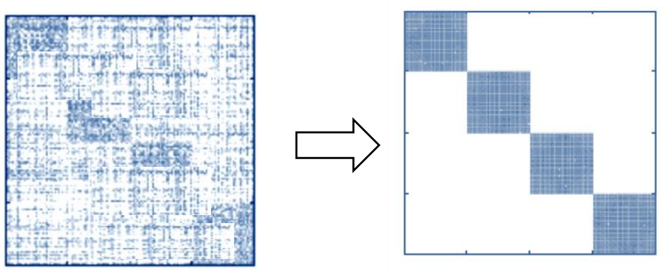
- SSC w/ block-diagonal prior
- LRR w/ block-diagonal

$$\min_{Z} \|Z\|_1 + \frac{\lambda}{2}\|X - XZ\|_F^2$$
$$\text{s.t. } \operatorname{diag}(Z) = 0\,,\ Z \in \mathcal{K}$$

BD-SSC

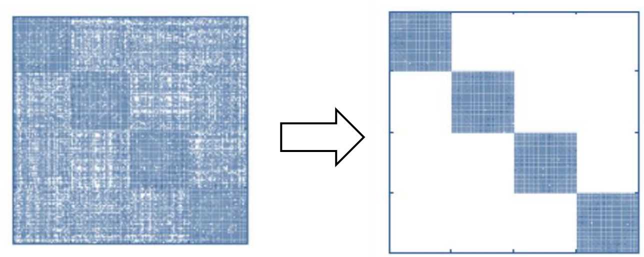
$$\min_{Z} \|Z\|_* + \frac{\lambda}{2}\|X - XZ\|_F^2$$
$$\text{s.t.}\, Z \in \mathcal{K}$$

BD-LRR



sparse affinity matrix

block-diagonal affinity matrix



low-rank affinity matrix

block-diagonal affinity matrix

Elhamifar, et al. *Sparse Subspace Clustering*. CVPR 2009

Liu, et al. *Robust Recovery of Subspace Structures by Low-Rank Representation*. TPAMI 2013

Elhamifar, et al. *Sparse Subspace Clustering: Algorithm, Theory, and Applications*. TPAMI 2013

Optimization

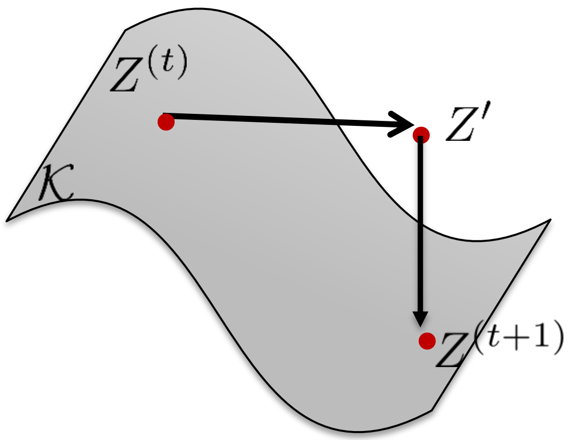
- Projected gradient descent for BD-SSC

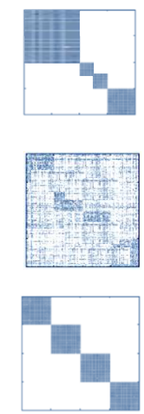
$$\min_{Z} \|Z\|_1 + \frac{\lambda}{2}\|X - XZ\|_F^2 \text{ s.t. } \mathrm{diag}(Z) = 0, Z \in \mathcal{K}$$

while not converged **do**

1. gradient $g^{(t)} \leftarrow \mathcal{G}(f(Z^{(t)}))YY^\top$;

end





project back:

$$\Pi_{\mathcal{K}} : Z' \to Z = \arg\min_{Z \in \mathcal{K}} \|Z - Z'\|_F^2$$

Feng, et al. *Robust Subspace Segmentation with Block-Diagonal Prior*. CVPR 2014

Convergence Guarantee

- Convergence
  - k -block diagonal constraint is non-convex
  - Converge to the global optimum with high probability if Scalable Restricted Isometry Property (SRIP) holds.

$\mathbf{SRIP}(d, \alpha)$: There exist $v_d, \mu_d > 0$ satisfying $\frac{\mu_d}{v_d} < \alpha$ such that:
$v_d \|\mathbf{x}\| \leq \|\mathcal{A}(\mathbf{x})\| \leq \mu_d \|\mathbf{x}\|$, for every $\mathbf{x} \in \{\mathbf{x} | \mathbf{x} \in \mathbb{E}, \phi(\mathbf{x}) \leq d\}$.

Theorem Consider the projective gradient algorithm with a constant step size $\eta_t = \eta \in [\mu_k^2, 2\nu_k^2)$ and suppose that $\mathrm{SRIP}(k, \sqrt{2})$ is satisfied by data. Then

$$f(Z^{(t+1)}) - f(Z^*) \leq (\rho - 1/2)^t \left\{ f(Z^{(0)}) - f(Z^*) \right\}, \forall t \geq 0$$

with $\rho = \eta / 2\nu_k^2$. As a consequence,

$$f(Z^{(t)}) \to f(Z^*) \text{ as } t \to \infty.$$

  - Convergence rate: $O(1/t)$

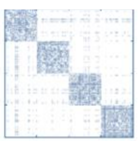
Beck and Teboulle. *A linearly convergent algorithm for solving a class of nonconvex/affine feasibility problems*. In Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, 2011.
Feng, et al. *Robust Subspace Segmentation with Block-Diagonal Prior*. CVPR 2014

Robustness?

- 5 sets of samples with different noise levels

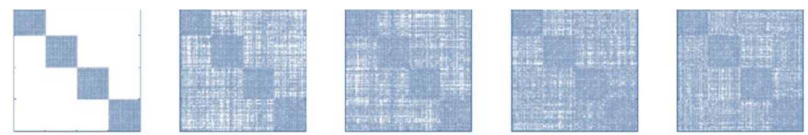
noise increasing

acc = 1.00



SSC

1.00 0.77 0.83 0.80 0.80



LRR

After re-arrangement, the matrices are block-diagonal.

Motion Segmentation

- Dataset: Hopkins 155
- Results (error rate %)

	SSC	BD-SSC	LRR	BD-LRR
Max	42.34	42.34	41.18	38.97
Mean	3.11	2.90	4.83	3.35
Med	0	0	0.52	0.37
Std	7.78	7.48	9.35	8.84

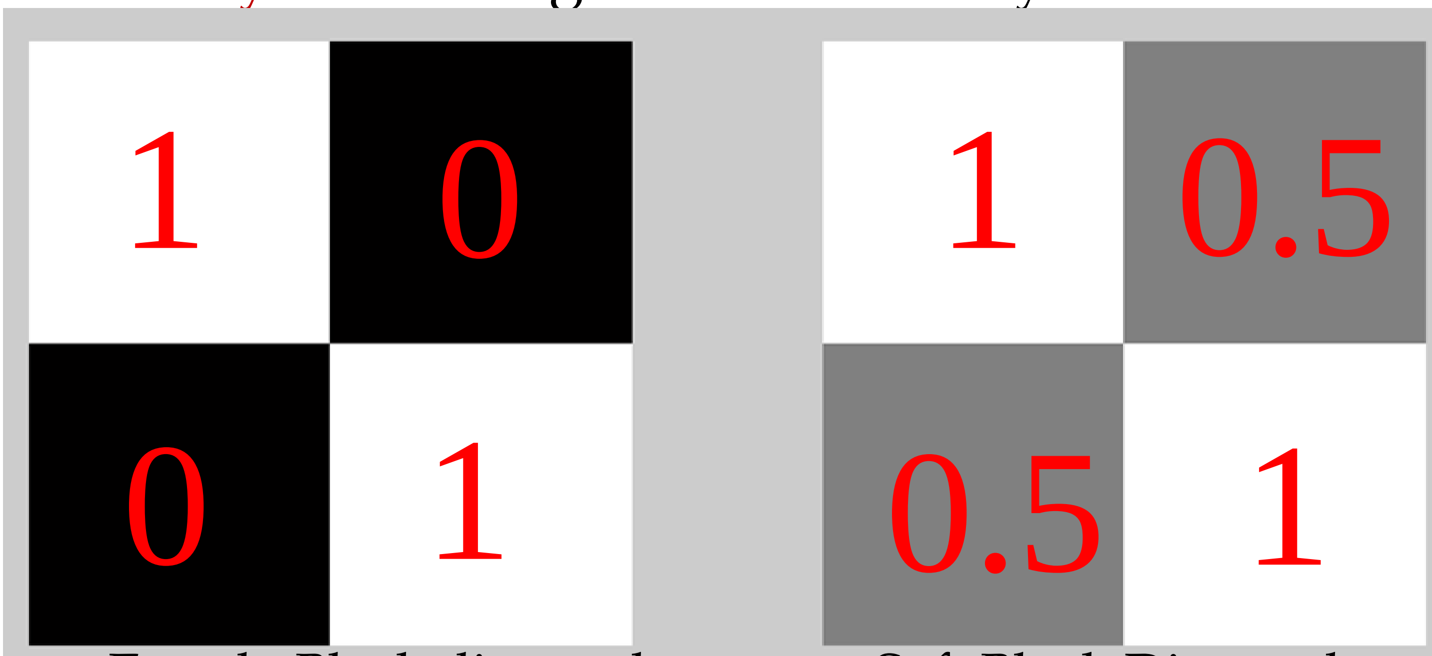
w/o post processing

	SSC	BD-SSC	LRR	BD-LRR
Max	47.20	43.07	43.38	14.93
Mean	1.99	1.68	1.74	0.97
Med	0	0	0	0
Std	6.67	5.97	5.51	2.46

w/ post processing

Exactly Block-diagonal Necessary?

- An exactly block diagonal matrix may not be necessary.



Exactly Block diagonal Soft Block Diagonal

- Both the above two affinity matrices lead to the same clustering results.

Soft Block Diagonal Representation

- A soft block diagonal representation model

Noiseless case

$$\min_{Z,B} \frac{1}{2}\|Z - B\|_F$$
$$\text{s.t. } X = XZ, \ B \text{ is } k \text{ block diagonal.}$$

Noisy case

$$\min_{Z,B} \frac{1}{2}\|X - XZ\|_F^2 + \frac{\lambda}{2}\|Z - B\|_F^2$$
$$\text{s.t. } B \text{ is } k \text{ block diagonal.}$$

- The above problems are nonconvex, difficult to solve.

Lu, et al. *Convex Sparse Spectral Clustering: Single-view to Multi-view*. TIP 2016.

Soft Block Diagonal Representation

- Convex Block Diagonal Representation (BDR)

$$\min_{Z,B} \frac{1}{2}\|X - XZ\|_F^2 + \frac{\lambda}{2}\|Z - B\|_F^2$$

$$\text{s.t. } B \in S_1 = \{B|\text{diag}(B) = 0, B \geq 0, B = B^\top\}$$

$$L_B \in S_2 = \{L|\mathbf{0} \preceq L \preceq \alpha I, \text{Tr}(L) = n - k.\}$$

Convex relaxation

B is k block diagonal

- There exists $\alpha > 1$ such that:

  $\text{rank}(L) = n - k$ i.e. B is k block diagonal

- Limitations: difficult to find proper α giving $\text{rank}(L) = n - k$
- In practice, $\alpha = 1.1$ works well.

Lu, et al. *Convex Sparse Spectral Clustering: Single-view to Multi-view*. TIP 2016.

Conclusions

- Strong error and outlier resistance.
- Theoretical guarantees and closed-form solutions.
- Block-diagonality plays a critical role in the performance of SC.
- Many interesting applications.

References

- Bin Cheng, Jianchao Yang, Shuicheng Yan, Yun Fu, Thomas S. Huang. Learning with l1-graph for Image Analysis. TIP 2010
- Ehsan Elhamifar, Rene Vidal. Sparse Subspace Clustering. CVPR 2009
- Ehsan Elhamifar, Rene Vidal. Sparse Subspace Clustering: Algorithm, Theory, and Applications. TPAMI 2013
- Guangcan Liu, Zhouchen Lin, Yong Yu. Robust Subspace Segmentation by Low-Rank Representation. ICML 2010
- Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, Yi Ma. Robust Recovery of Subspace Structures by Low-Rank Representation. TPAMI 2013
- Dijun Luo, Feiping Nie, Chris Ding, Heng Huang. Multi-Subspace Representation and Discovery, ECML PKDD 2011
- Shusen Wang, Xiaotong Yuan, Tiansheng Yao, Shuicheng Yan, Jialie Shen. Efficient Subspace Segmentation via Quadratic Programming, AAAI 2011
- Canyi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, Shuicheng Yan. Robust and Efficient Subspace Segmentation via Least Squares Regression. ECCV 2012
- Han Hu, Zhouchen Lin, Jianjiang Feng, Jie Zhou. Smooth Representation Clustering. CVPR 2014
- Jiashi Feng, Zhouchen Lin, Huan Xu, Shuicheng Yan. Robust Subspace Segmentation with Block-Diagonal Prior. CVPR 2014
- Canyi Lu, Zhouchen Lin, Shuicheng Yan. Convex Sparse Spectral Clustering: Single-view to Multi-view. TIP 2016.

Afternoon Session

- Afternoon Session: 2PM

- 2:00 – 3:00 PM Algorithms & More Models
- 3:00 – 3:30 PM Coffee Break
- 3:30 – 5:00 PM Applications