

Low Rank Representation - Theories, Algorithms, Applications

林宙辰

北京大学

May 3, 2012

Outline

- Low Rank Representation
- Some Theoretical Analysis
- Solution by LADM
- Applications
- Generalizations
- Conclusions

Sparse Representation

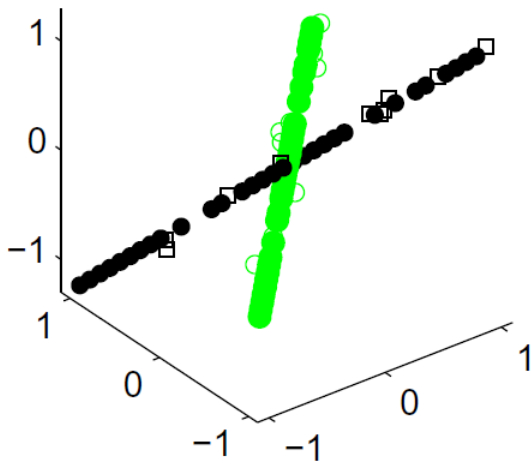
- Sparse Representation

$$\begin{aligned} \min \|x\|_0, \\ \text{s.t. } y = Ax. \end{aligned} \tag{1}$$

- Sparse Subspace Clustering

$$\begin{aligned} \min \|z_i\|_0, \\ \text{s.t. } x_i = X_{\hat{i}} z_i, \quad \forall i. \end{aligned} \tag{2}$$

where $X_{\hat{i}} = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$.



$$\begin{aligned} \min \|Z\|_0, \\ \text{s.t. } X = XZ, \text{diag}(Z) = 0. \end{aligned} \tag{3}$$

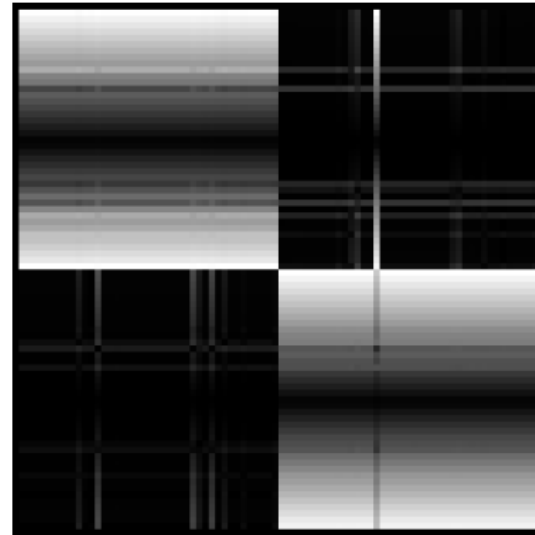
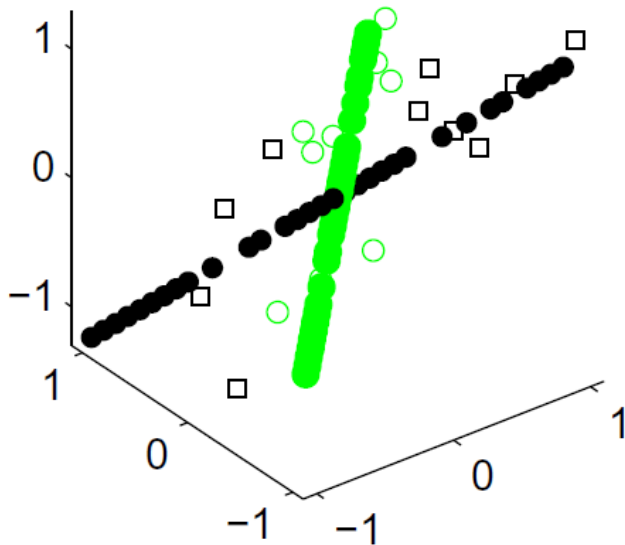
$$\begin{aligned} \min \|Z\|_1, \\ \text{s.t. } X = XZ, \text{diag}(Z) = 0. \end{aligned} \tag{4}$$

Sparse Representation

- Construct a graph

$$W = (|Z^*| + |(Z^*)^T|)/2$$

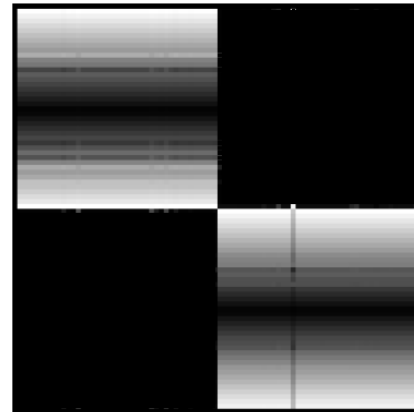
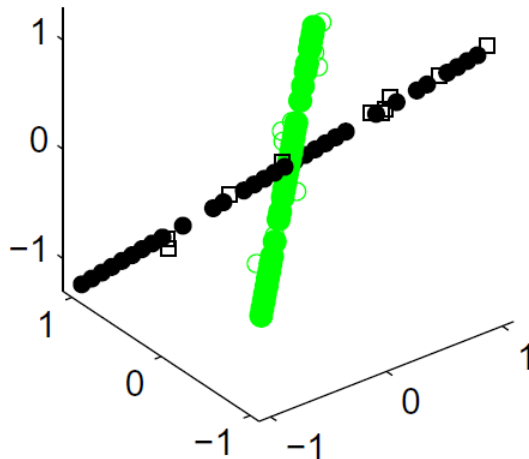
- Normalized cut on the graph



Sparse Representation

Theorem. Assume the data is clean and is drawn from independent subspaces, then Z^* is block diagonal.

$$\dim\left(\sum_i S_i\right) = \sum_i \dim(S_i).$$



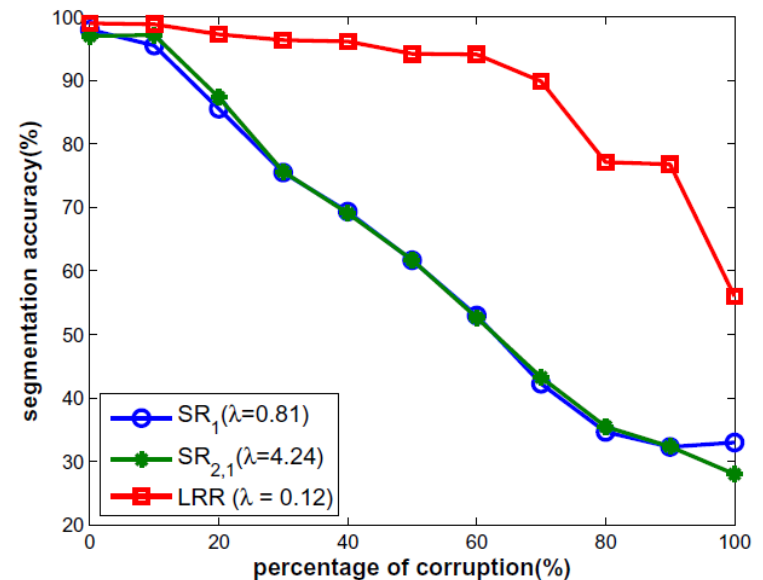
Drawback of SSC

- Sensitive to noise: no cross validation among coefficients

$$\begin{aligned} \min & \|Z\|_1, \\ \text{s.t.} & X = XZ, \text{diag}(Z) = 0. \end{aligned} \quad (4)$$



$$\begin{aligned} \min & \|z_i\|_1, \\ \text{s.t.} & x_i = Xz_i, (z_i)_i = 0. \end{aligned} \quad (5)$$



Hints from 2D Sparsity

- Rank is a good measure of 2D sparsity
 - Real data usually lie on low-dim manifolds



low-dim subspaces \rightarrow low rank data matrices

- Low rank \leftrightarrow high correlation among rows/columns

Low Rank Representation

$$\begin{aligned} \min \|Z\|_1, \\ \text{s.t. } X = XZ, \text{diag}(Z) = 0. \end{aligned} \tag{4}$$

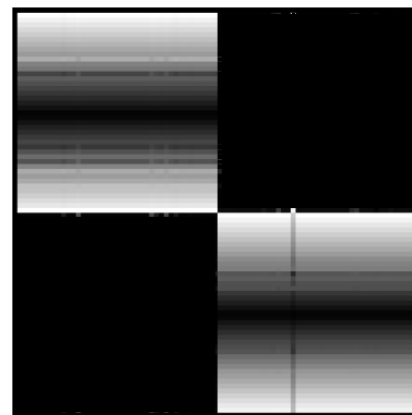
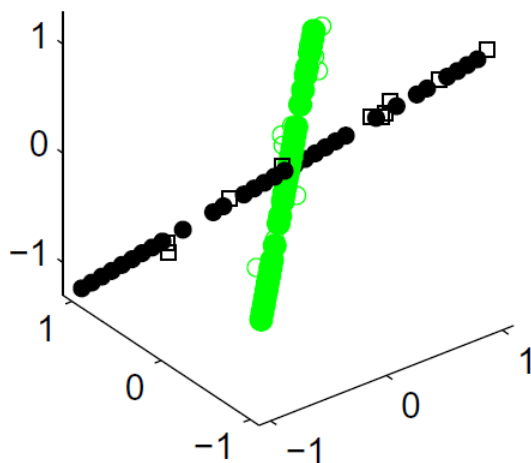
$$\begin{aligned} \min \|Z\|_*, \\ \text{s.t. } X = XZ. \end{aligned} \tag{6}$$

no additional
constraint!

$\|Z\|_* = \sum_j \sigma_j(Z)$, nuclear norm, a convex surrogate of rank.

Low Rank Representation

Theorem. Assume the data is clean and is drawn from independent subspaces, then there exists Z^* which is block diagonal, and the rank of each block equals the dimension of the corresponding subspace.



Low Rank Representation

- When there is noise and outliers

$$\begin{aligned} \min & \|Z\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.} & X = XZ + E. \end{aligned} \quad (7)$$

where $\|E\|_{2,1} = \sum_i \|E_{:,i}\|_2$.

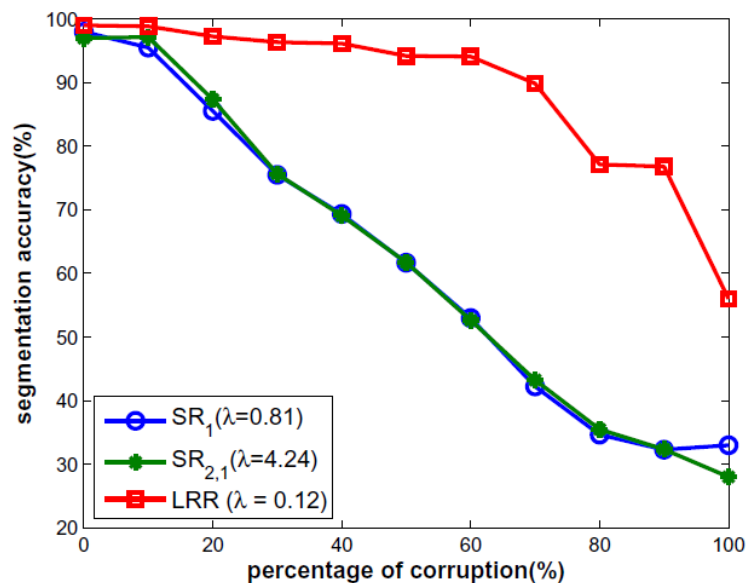


Table 2. Segmentation accuracy (%) on Extended Yale Database B. We have tuned the parameters of all methods to the best. The parameter of LRR is set to be $\lambda = 0.15$

	GPCA	LSA	RANSAC	SSC	LRR
Acc.	NA	31.72	NA	37.66	62.53

Low Rank Representation

- Connection to Robust PCA

$$\begin{aligned} \min_{A,E} & \|A\|_* + \lambda \|E\|_1, \\ \text{s.t.} & X = A + E, \end{aligned} \tag{8}$$

The clean data is low rank **w.r.t.** the dictionary I .

$$\begin{aligned} \min & \|Z\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.} & X = XZ + E. \end{aligned} \tag{7}$$

The clean data is low rank **w.r.t.** the dictionary X .

Low Rank Representation

- Generalization

$$\begin{aligned} \min & \|Z\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.} & X = AZ + E. \end{aligned} \tag{9}$$

The clean data is low rank **w.r.t.** the dictionary A .

More Theoretical Analysis

- Closed form solution at noiseless case

$$\begin{aligned} \min_Z & \|Z\|_*, \\ \text{s.t.} & X = XZ, \end{aligned}$$

has a *unique closed-form* optimal solution: $Z^* = V_r V_r^T$, where $U_r \Sigma_r V_r^T$ is the skinny SVD of X .

- Shape Interaction Matrix
- when X is sampled from independent subspaces, Z^* is block diagonal, each block corresponding to a subspace

$$\min_{X=XZ} \|Z\|_* = \text{rank}(X).$$

More Theoretical Analysis

- Closed form solution at general case

$$\min_Z \|Z\|_*, \quad s.t. \quad X = AZ,$$

has a *unique closed-form* optimal solution: $Z^* = A^\dagger X$.

Follow-up Work

- More problems with closed form solutions

$$\min_Z \varepsilon \|Z\|_* + \frac{1}{2} \|XZ - X\|_F^2$$

$$\min_{A,Z} \varepsilon \|Z\|_* + \frac{\tau}{2} \|AZ - A\|_F^2 + \frac{1}{2} \|D - A\|_F^2$$

$$\min_{A,Z} \varepsilon \|Z\|_* + \frac{1}{2} \|A - X\|_F^2, \quad s.t. \quad A = AZ.$$

- Speeding up optimization

Exact Recoverability of LRR

Theorem: Let $\lambda = 3/(7\|X\|\sqrt{\gamma n})$. Then there exists $\gamma^* > 0$ such that when $\gamma \leq \gamma^*$, LRR can exactly recover the row space and the column support of (Z_0, E_0) :

$$U^*(U^*)^T = V_0V_0^T, \quad \mathcal{I}^* = \mathcal{I}_0,$$

where $\gamma = |\mathcal{I}_0|/n$ is the fraction of outliers, U^* is the column space of Z^* , V_0^T is the row space of Z_0 , and \mathcal{I}^* and \mathcal{I}_0 is the column supports of E^* and E_0 , respectively.

Linearized Alternating Direction Method (LADM)

- Model Problem

$$\min_{\mathbf{x}, \mathbf{y}} f(\mathbf{x}) + g(\mathbf{y}), \text{ s.t. } \mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{y}) = \mathbf{c},$$

where \mathbf{x} , \mathbf{y} and \mathbf{c} could be either vectors or matrices, f and g are convex functions, and \mathcal{A} and \mathcal{B} are linear mappings.

- ADM

$$\mathcal{L}_A(\mathbf{x}, \mathbf{y}, \lambda) = f(\mathbf{x}) + g(\mathbf{y}) + \langle \lambda, \mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{y}) - \mathbf{c} \rangle + \frac{\beta}{2} \|\mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{y}) - \mathbf{c}\|^2,$$

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} \mathcal{L}_A(\mathbf{x}, \mathbf{y}_k, \lambda_k),$$

$$\mathbf{y}_{k+1} = \arg \min_{\mathbf{y}} \mathcal{L}_A(\mathbf{x}_{k+1}, \mathbf{y}, \lambda_k),$$

$$\lambda_{k+1} = \lambda_k + \beta[\mathcal{A}(\mathbf{x}_{k+1}) + \mathcal{B}(\mathbf{y}_{k+1}) - \mathbf{c}].$$

- Difficulties

Linearized Alternating Direction Method (LADM)

$$\begin{aligned}\mathbf{x}_{k+1} &= \arg \min_{\mathbf{x}} f(\mathbf{x}) + \frac{\beta}{2} \|\mathcal{A}(\mathbf{x}) + \mathcal{B}(\mathbf{y}_k) - \mathbf{c} + \lambda_k/\beta\|^2, \\ \mathbf{y}_{k+1} &= \arg \min_{\mathbf{y}} g(\mathbf{y}) + \frac{\beta}{2} \|\mathcal{A}(\mathbf{x}_{k+1}) + \mathcal{B}(\mathbf{y}) - \mathbf{c} + \lambda_k/\beta\|^2\end{aligned}$$

- Linearize the quadratic term

$$\begin{aligned}\mathbf{x}_{k+1} &= \arg \min_{\mathbf{x}} f(\mathbf{x}) + \langle \mathcal{A}^*(\lambda_k) + \beta \mathcal{A}^*(\mathcal{A}(\mathbf{x}_k) + \mathcal{B}(\mathbf{y}_k) - \mathbf{c}), \mathbf{x} - \mathbf{x}_k \rangle + \frac{\beta \eta_A}{2} \|\mathbf{x} - \mathbf{x}_k\|^2 \\ &= \arg \min_{\mathbf{x}} f(\mathbf{x}) + \frac{\beta \eta_A}{2} \|\mathbf{x} - \mathbf{x}_k + \mathcal{A}^*(\lambda_k + \beta(\mathcal{A}(\mathbf{x}_k) + \mathcal{B}(\mathbf{y}_k) - \mathbf{c})) / (\beta \eta_A)\|^2,\end{aligned}$$

$$\mathbf{y}_{k+1} = \arg \min_{\mathbf{y}} g(\mathbf{y}) + \frac{\beta \eta_B}{2} \|\mathbf{y} - \mathbf{y}_k + \mathcal{B}^*(\lambda_k + \beta(\mathcal{A}(\mathbf{x}_{k+1}) + \mathcal{B}(\mathbf{y}_k) - \mathbf{c})) / (\beta \eta_B)\|^2.$$

- Adaptive Penalty

Linearized Alternating Direction Method (LADM)

Theorem: If $\{\beta_k\}$ is non-decreasing and upper bounded, $\eta_A > \|\mathcal{A}\|^2$, and $\eta_B > \|\mathcal{B}\|^2$, then the sequence $\{(\mathbf{x}_k, \mathbf{y}_k, \lambda_k)\}$ generated by LADMAP converges to a KKT point of the model problem.

Applying LADM to LRR

- LRR

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

- Further Acceleration Technique
 - $O(n^3) \rightarrow O(rn^2)$

Experiments

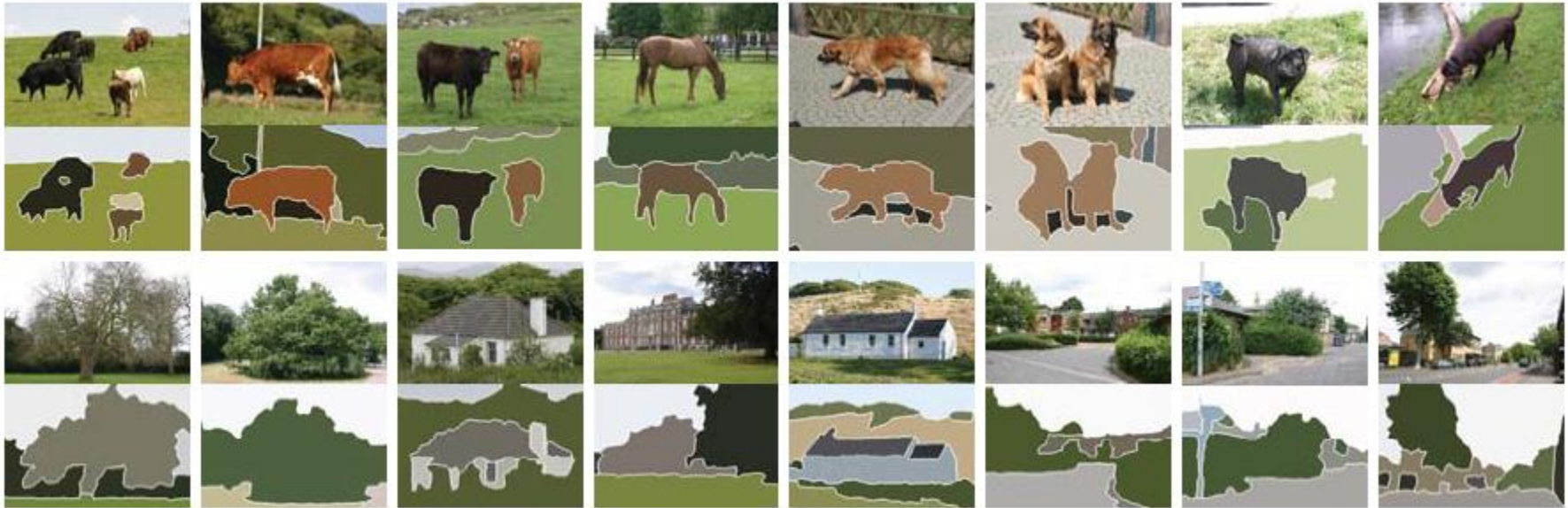
Table 1: Comparison among APG, ADM, LADM, standard LADMAP and accelerated LADMAP (denoted as LADMAP(A)) on the synthetic data. For each quadruple (s, p, d, \tilde{r}) , the LRR problem, with $\mu = 0.1$, was solved for the same data using different algorithms. We present typical running time (in $\times 10^3$ seconds), iteration number, relative error (%) of output solution $(\hat{\mathbf{E}}, \hat{\mathbf{Z}})$ and the clustering accuracy (%) of tested algorithms, respectively.

Size (s, p, d, \tilde{r})	Method	Time	Iter.	$\frac{\ \mathbf{Z}-\mathbf{Z}_0\ }{\ \mathbf{Z}_0\ }$	$\frac{\ \mathbf{E}-\mathbf{E}_0\ }{\ \mathbf{E}_0\ }$	Acc.
(10, 20,200, 5)	APG	0.0332	110	2.2079	1.5096	81.5
	ADM	0.0529	176	0.5491	0.5093	90.0
	LADM	0.0603	194	0.5480	0.5024	90.0
	LADMAP	0.0145	46	0.5480	0.5024	90.0
	LADMAP(A)	0.0010	46	0.5480	0.5024	90.0
(15, 20,300, 5)	APG	0.0869	106	2.4824	1.0341	80.0
	ADM	0.1526	185	0.6519	0.4078	83.7
	LADM	0.2943	363	0.6518	0.4076	86.7
	LADMAP	0.0336	41	0.6518	0.4076	86.7
	LADMAP(A)	0.0015	41	0.6518	0.4076	86.7
(20, 25, 500, 5)	APG	1.8837	117	2.8905	2.4017	72.4
	ADM	3.7139	225	1.1191	1.0170	80.0
	LADM	8.1574	508	0.6379	0.4268	80.0
	LADMAP	0.7762	40	0.6379	0.4268	84.6
	LADMAP(A)	0.0053	40	0.6379	0.4268	84.6
(30, 30, 900, 5)	APG	6.1252	116	3.0667	0.9199	69.4
	ADM	11.7185	220	0.6865	0.4866	76.0
	LADM	N.A.	N.A.	N.A.	N.A.	N.A.
	LADMAP	2.3891	44	0.6864	0.4294	80.1
	LADMAP(A)	0.0058	44	0.6864	0.4294	80.1

Applications of LRR

- Image segmentation

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

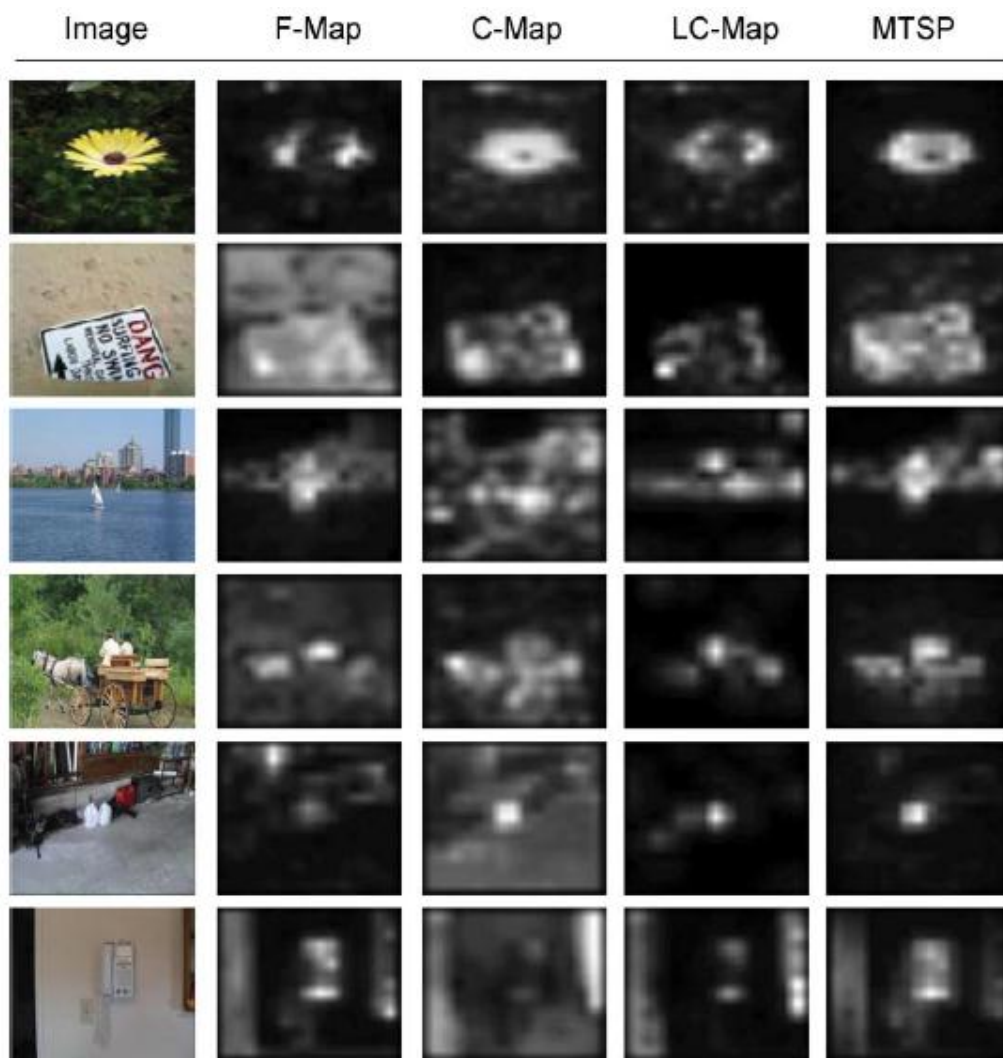
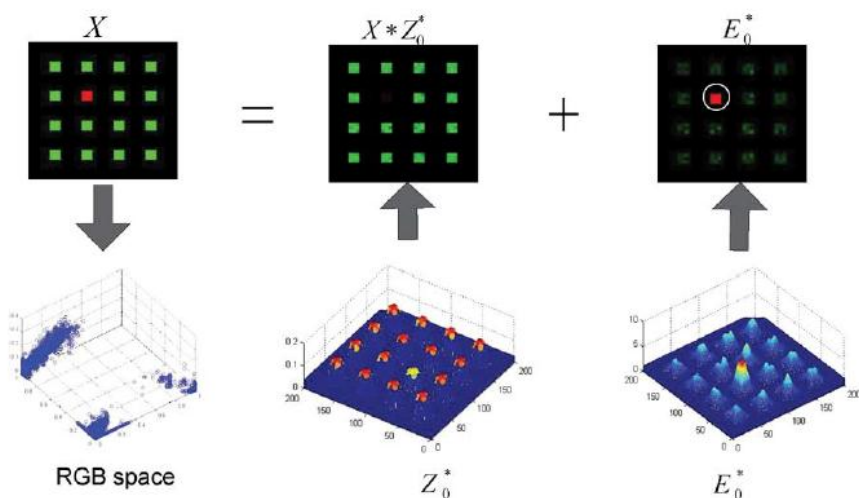


Applications of LRR

- Saliency detection

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_{2,1},$$

$$s.t. \quad X = XZ + E.$$



Generalizations of LRR

- LRR with clean data

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

TABLE I
SEGMENTATION ERROR RATE (%) ON THE HOPKIN155 DATABASE. THE
PARAMETERS ARE SELECTED AS $\lambda_{LRR} = 2.4$ AND $\lambda_{RSI} = 0.24$,
RESPECTIVELY.

METHOD	MEAN	MEDIAN	STD
LRR	4.3673	0.4717	7.4540
RSI	2.8501	0	7.5858

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

Table 1. Errors on Hopkins155 data without pre/post-processing.

Method	LSA	SIM	SSC	LRR	OUR
Average	8.99%	5.25%	3.89%	3.16%	3.28%

Generalizations of LRR

- Latent LRR
 - To address insufficient sample problem

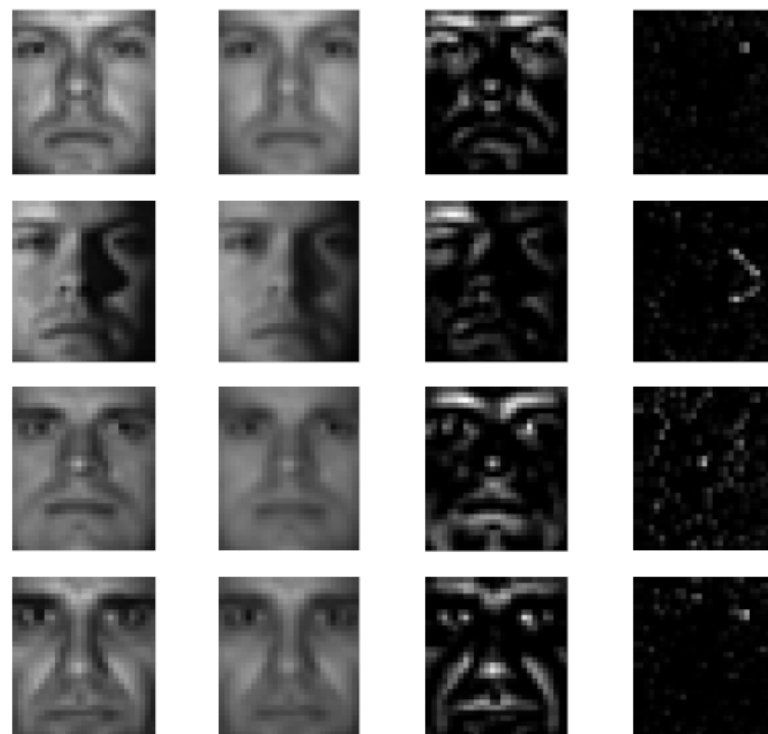
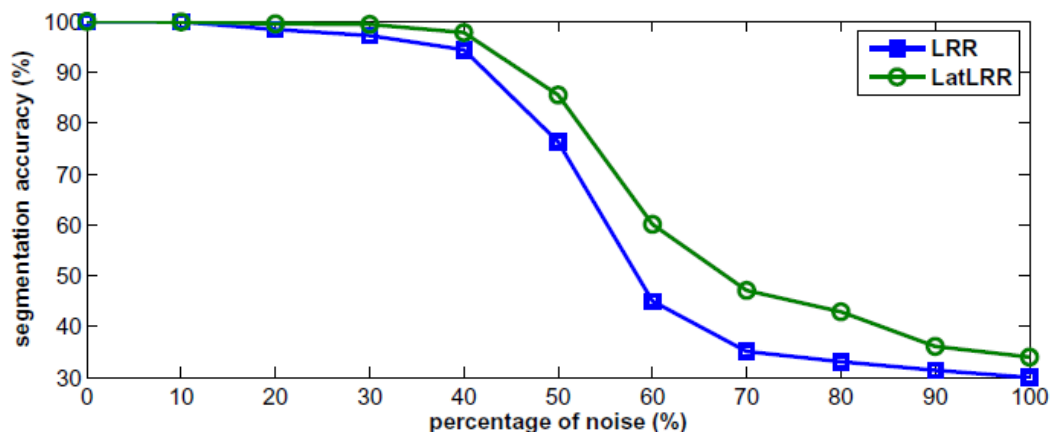
$$(X = [X, X_H]Z_{O,H}^* \implies X = XZ + LX)$$

$$\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda\|E\|_{2,1},$$

$$s.t. \quad X = XZ + LX + E.$$

$$X = XZ^* + L^*X + E^*$$

data = principal features + salient features + sparse noise



Generalizations of LRR

- Fixed Rank Representation (FRR)
 - To address insufficient sample problem
 - Subspace clustering: using \tilde{Z}^*

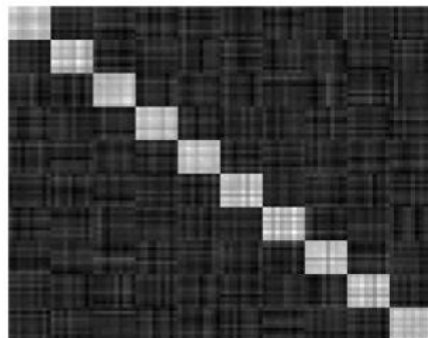
$$\min_{Z, \tilde{Z}, E} \|Z - \tilde{Z}\|_F^2 + \lambda \|E\|_{2,1}, \quad s.t. \quad X = XZ + E, \text{rank}(\tilde{Z}) \leq r.$$

- Feature extraction: by $y = \tilde{Z}^* x$

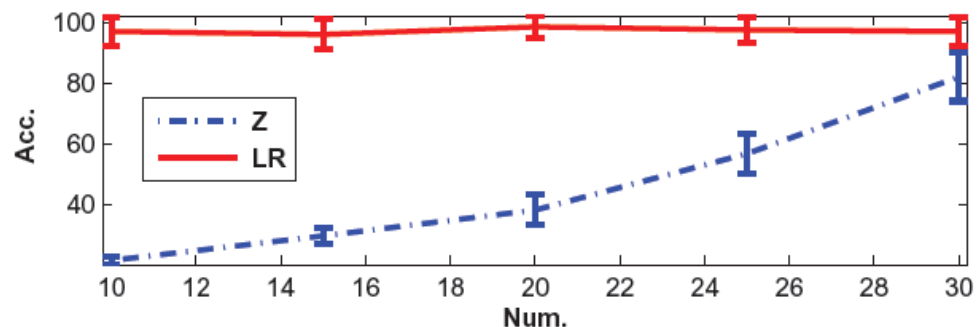
$$\min_{Z, \tilde{Z}, E} \|Z - \tilde{Z}\|_F^2 + \lambda \|E\|_{2,1}, \quad s.t. \quad X = ZX + E, \text{rank}(\tilde{Z}) \leq r.$$



(a) Z



(b) LR



Generalizations of LRR

- Semi-supervised learning

$$\min_{Z, E} \|Z\|_* + \beta \|Z\|_1 + \lambda \|E\|_{2,1}, \quad s.t. \quad X = AZ + E, Z \geq 0.$$

propagate labels on the graph with weights $(Z^* + (Z^*)^T)/2$.

Dataset	k NN0	k NN1	LLE0	LLE1	ℓ_1 -graph	SPG	LRR-graph	NNLRS-graph
YaleB (10%)	33.51	38.27	29.21	29.94	46.13	15.57	28.22	3.75
YaleB (20%)	34.66	38.97	30.63	30.63	45.54	17.56	24.46	9.84
YaleB (30%)	33.71	37.87	28.17	28.17	46.14	16.54	22.33	10.54
YaleB (40%)	33.00	37.34	28.36	28.36	43.39	17.16	19.42	9.38
YaleB (50%)	33.10	37.38	28.38	28.38	42.25	18.99	18.04	9.64
YaleB (60%)	32.48	37.78	28.53	28.53	41.52	20.50	16.09	8.13
PIE (10%)	34.84	37.54	33.06	33.44	22.88	20.50	33.98	11.11
PIE (20%)	37.46	40.31	35.05	35.81	22.94	20.30	34.35	22.81
PIE (30%)	35.30	37.80	32.52	32.88	22.33	20.60	31.81	17.86
PIE (40%)	35.81	38.22	32.51	32.99	23.14	20.81	32.39	16.25
PIE (50%)	34.39	37.38	31.41	31.64	23.01	21.43	31.33	19.25
PIE (60%)	35.63	38.00	32.76	32.85	25.76	23.82	32.50	21.56
USPS (10%)	1.87	2.20	17.10	27.31	43.27	3.95	2.25	1.57
USPS (20%)	2.51	2.67	22.92	30.83	41.27	5.28	3.10	1.93
USPS (30%)	5.88	6.10	21.26	27.54	38.31	10.48	8.91	4.95
USPS (40%)	7.87	8.44	19.21	22.78	34.86	14.22	13.44	7.44
USPS (50%)	17.19	18.44	18.41	19.48	29.42	20.38	21.88	11.27
USPS (60%)	11.04	15.20	14.80	14.94	23.36	15.89	17.75	6.09

Generalizations of LRR

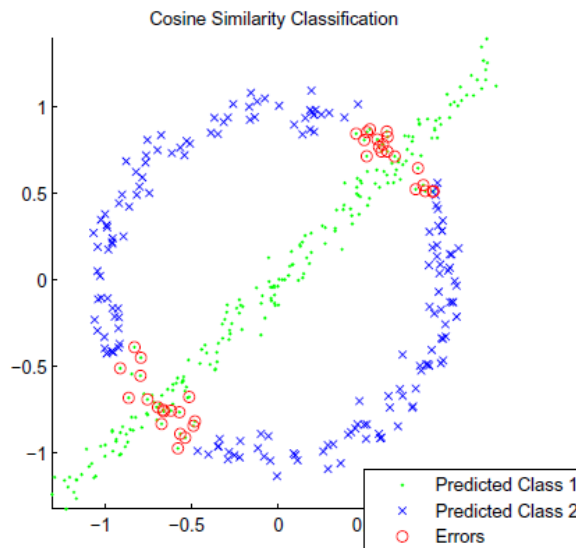
- Kernel LRR
 - To address nonlinear multi-manifold segmentation

$$\min_Z \|\phi(X) - \phi(X)Z\|_F^2 + \lambda \|Z\|_*.$$



Kernel trick: $\langle \phi(x), \phi(y) \rangle = K(x, y)$

$$\min_Z \sum_i (z_i^T K(X, X) z_i - 2K(x_i, X) z_i + K(x_i, x_i)) + \lambda \|Z\|_*.$$



Thanks!

- zlin@pku.edu.cn

$$\begin{aligned} \min_{Z, E} & \|Z\|_* + \lambda \|E\|_{2,1}, \\ \text{s.t.} & X = XZ + E. \end{aligned}$$

LRR = SSC + RPCA

