Graph based Subspace Segmentation

Canyi Lu

National University of Singapore

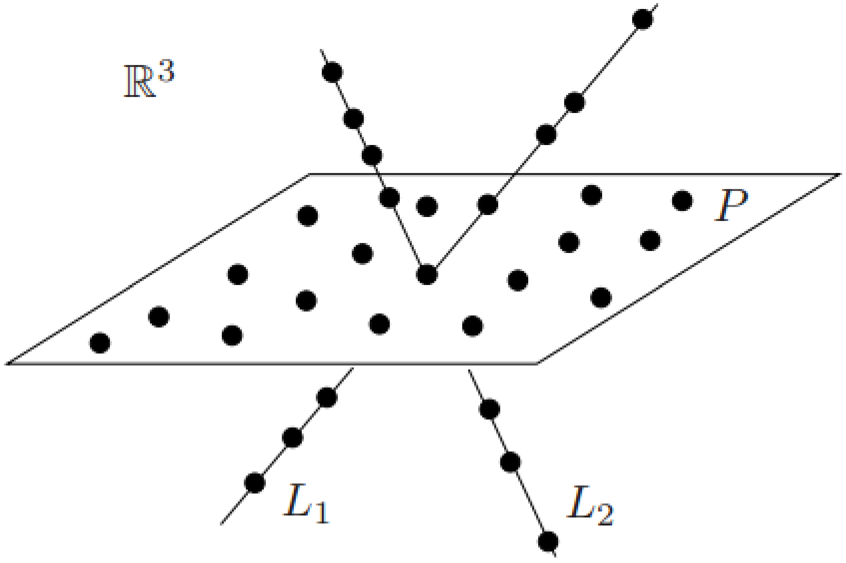
Nov. 21, 2013

Content

- Subspace Segmentation Problem
- Related Work
  - Sparse Subspace Clustering (SSC)
  - Low-Rank Representation (LRR)
  - Multi-Subspace Representation (MSR)
  - Subspace Segmentation via Quadratic Programming (SSQP)
  - Least Squares Regression (LSR)
- Enforced Block Diagonal Conditions
- Correlation Adaptive Subspace Segmentation (CASS)

Subspace Segmentation

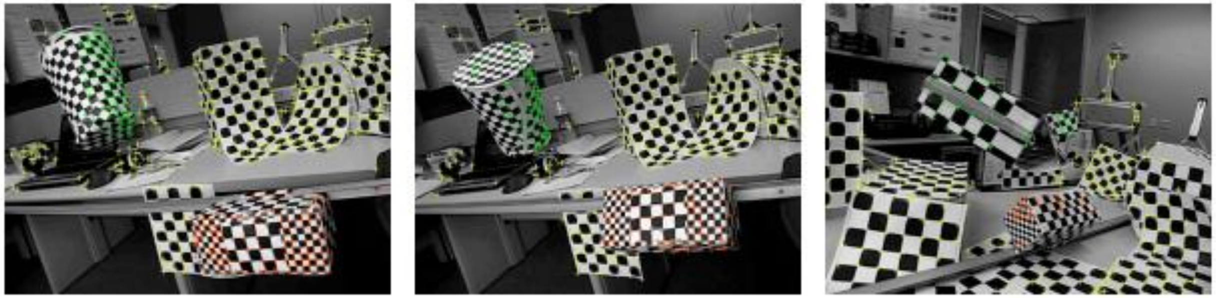
- Given sufficient data points drawn from multiple subspaces, the goal is to find
  - the number of subspaces
  - their dimensions
  - a basis of each subspace
  - the segmentation of the data corresponding to different subspaces



Applications

- Face clustering

- Motion segmentation

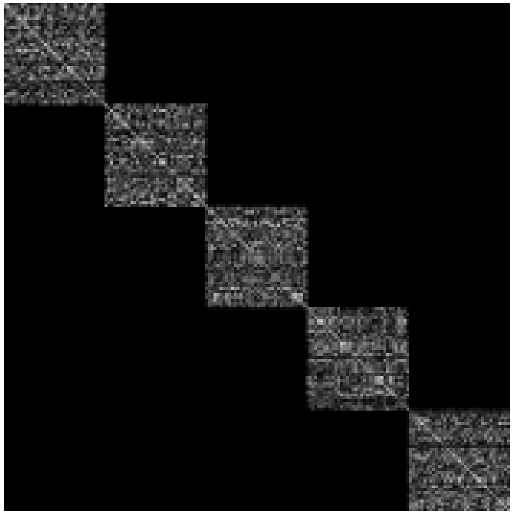


- Image segmentation

Algorithm

Spectral Clustering：

- Graph construction: construct a graph (affinity matrix) to measure the similarities between data points



- Segment the data points into multiple clusters

Subspace Assumption

- Disjoint subspaces

Two subspaces are said to be disjoint if they intersect only at the orgin.
k subspaces $\{\mathcal{S}_i\}_{i=1}^{k}$ are said to be disjoint if every two subspaces are disjoint.

- Independent subspaces

k subspaces $\{\mathcal{S}_i\}_{i=1}^{k}$ are said to be independent if $\dim(\oplus_{i=1}^{k}\mathcal{S}_i)=\sum_{i=1}^{k}\dim(\mathcal{S}_i)$, where $\oplus$ is the direct sum.

- Orthogonal subspaces

Previous Work: SSC

- Graph construction by sparse representation

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

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

- The solution to the above L1 minimization problem is block diagonal when the data are from independent subspace.

Bin Cheng, Jianchao Yang, Shuicheng Yan, Yun Fu, Thomas S. Huang, Learning with l1-graph for image analysis. TIP, 2010
Elhamifar, E. and R. Vidal. Sparse Subspace Clustering. CVPR 2009

Previous Work: LRR

- Graph construction by low rank representation

$$\min \operatorname{rank}(Z) \;\; \text{s.t. } X = XZ,$$

$$\min \|Z\|_* \;\; \text{s.t. } X = XZ,$$

- The solution to the above nuclear norm minimization problem is block diagonal when the data are from independent subspace.

Liu, G., Z. Lin, S. Yan, J. Sun, Y. Yu and Y. Ma. Robust recovery of subspace structures by low-rank representation. TPAMI. 2013

Previous Work: MSR

- Graph construction by low rank representation

$$\min ||Z||_* + \delta ||Z||_1 \;\; \text{s.t. } X = XZ, \text{diag}(Z) = 0.$$

- The solution to the above minimization problem is block diagonal when the data are from independent subspace.

Previous Work: SSQP

- Subspace Segmentation via Quadratic Programming

$$\min ||XZ - X||_F^2 + \lambda ||Z^T Z||_1 \ \text{ s.t. } Z \geq 0, \text{diag}(Z) = 0.$$

- The solution to the above nuclear norm minimization is block diagonal when the data are from orthogonal subspace.

Least Squares Regression

- Subspace Segmentation via Least Squares Regression

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

- The solution to the above minimization is block diagonal when the data are from independent subspace.
- Grouping effect of LSR (in vector form)

$$\min ||y - Xz||_2^2 + \lambda ||z||_2^2.$$

We have

$$\frac{||z_i^* - z_j^*||_2}{||y||_2} \leq \frac{1}{\lambda}\sqrt{2(1-r)},$$

where $r = x_i^T x_j$ *is the sample correlation.*

Canyi Lu, Hai Min, Shuicheng Yan. Efficient Subspace Segmentation via Least Squares Regression. ECCV. 2012.

- Consider the following general problem

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

- What kind of objective function involves the block diagonal property under certain condition?

Enforced Block Diagonal Conditions

Enforced Block Diagonal Conditions A function f is defined on $\Omega(\neq \varnothing)$ which is a set of matrices. 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 $Z^D = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \in \Omega$. We require

(1) $f(Z) = f(P^T Z P)$, for any permutation matrix P, $P^T Z P \in \Omega$.
(2) $f(Z) \geq f(Z^D)$, where the equality holds if and only if $B = C = 0$ (or $Z = Z^D$).
(3) $f(Z^D) = f(A) + f(D)$.

Block Diagonal Property: independent subspaces

Theorem 1. *Assume the data sampling is sufficient, and the subspaces are independent. If f satisfies the EBD conditions (1)(2), the optimal solution Z^* to the problem*

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

is block diagonal:

$$Z^* = \begin{bmatrix} Z_1^* & 0 & \cdots & 0 \\ 0 & Z_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_k^* \end{bmatrix}$$

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 also the optimal solution to the following problem:

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

Canyi Lu, Hai Min, Shuicheng Yan. Efficient Subspace Segmentation via Least Squares Regression. ECCV. 2012.

Block Diagonal Property

Table 1. Criteria which satisfy the EBD conditions (1)(2)(3).

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

Proposition 1 *If f_i satisfies the EBD conditions (1)(2)(3) on Ω_i, then also $\sum_{i=1}^m \lambda_i f_i$, $(\lambda_i > 0)$ on $\cap_{i=1}^m \Omega_i (\neq \varnothing)$.*

Block Diagonal Property : orthogonal subspaces

Block diagonal property when the subspaces are orthogonal

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

$$\min ||X - XZ||_{2,p} + \lambda f(Z) \tag{9}$$

must be block diagonal, where $||\cdot||_{2,p}$ is defined as $||M||_{2,p} = (\sum_j (\sum_{i=1}^n M_{ij}^2)^{\frac{p}{2}})^{\frac{1}{p}}$, $p > 0$, and $\lambda > 0$ is a parameter which balances the effects of two terms.

Block Diagonal Property : disjoint subspaces

- Block diagonal property by SSC on disjoint subspaces

Theorem 3. *Assume the data points are sampled from k subspaces $\{\mathcal{S}_i\}_{i=1}^k$ of dimensions $\{d_i\}_{i=1}^n$. Let X_i denote the data points on $\mathcal{S}_i$ and $\hat{X}_i$ denote the data points on the other subspaces. Let $\mathbb{W}_i$ be the set of all full rank submatrices $\check{X}_i \in \mathbb{R}^{D \times d_i}$ of X_i. If the sufficient condition*

$$\max_{\check{X}_i \in \mathbb{W}_i} \sigma_{d_i}(\check{X}_i) > \sqrt{d_i} \Delta_i \max_{j \neq i} \cos(\theta_{ij})$$

is satisfied for all $i \in \{1, \cdots, k\}$, then for every nonzero $y \in \mathcal{S}_i$, the solution to the following problem

$$\begin{bmatrix} c_i^* \\ \hat{c}_i^* \end{bmatrix} = \arg\min \left\| \begin{bmatrix} c_i \\ \hat{c}_i \end{bmatrix} \right\|_1 \quad s.t. \ y = [X_i, \hat{X}_i] \begin{bmatrix} c_i \\ \hat{c}_i \end{bmatrix}.$$

Block Diagonal Property : disjoint subspaces

- Block diagonal property by LSR on disjoint subspaces

Theorem 4. *Assume the data points are sampled from k subspaces $\{\mathcal{S}_i\}_{i=1}^k$ of dimensions $\{d_i\}_{i=1}^n$. Let X_i denote the data points on $\mathcal{S}_i$ and $\hat{X}_i$ denote the data points on the other subspaces. Let $\mathbb{W}_i$ be the set of all full rank submatrices $\check{X}_i \in \mathbb{R}^{D \times d_i}$ of X_i. If the sufficient condition*

$$\max_{\check{X}_i \in \mathbb{W}_i} \sigma_{d_i}(\check{X}_i) > \sqrt{D}\Delta_i \max_{j \neq i} \cos(\theta_{ij})$$

is satisfied for all $i \in \{1, \cdots, k\}$, then for every nonzero $y \in \mathcal{S}_i$, the solution to the following problem

$$\begin{bmatrix} c_i^* \\ \hat{c}_i^* \end{bmatrix} = \arg\min \left|\left| \begin{bmatrix} c_i \\ \hat{c}_i \end{bmatrix} \right|\right|_2 \quad s.t.\ y = [X_i, \hat{X}_i] \begin{bmatrix} c_i \\ \hat{c}_i \end{bmatrix}.$$

Correlation Adaptive Subspace Segmentation

- A better choice: balance the sparsity and grouping effect
- Correlation Adaptive Subspace Segmentation (CASS)

$$\min_{w \in \mathbb{R}^n} ||X\mathrm{Diag}(w)||_* \ \text{ s.t. } \ y = Xw.$$

- If the data are uncorrelated (the data points are orthogonal $X^T X = I$)

$$||X\mathrm{Diag}(w)||_* = ||\mathrm{Diag}(w)||_* = \sum_{i=1}^{n} |w_i| = ||w||_1.$$

- If the data are highly correlated (the data points are all the same, $X = x_1 \mathbf{1}^T$, $X^T X = \mathbf{1}\mathbf{1}^T$)

$$||X\mathrm{Diag}(w)||_* = ||x_1 w^T||_* = ||x_1||_2 ||w||_2 = ||w||_2.$$

- For other case,

$$||w||_2 \leq ||X\mathrm{Diag}(w)||_* \leq ||w||_1$$

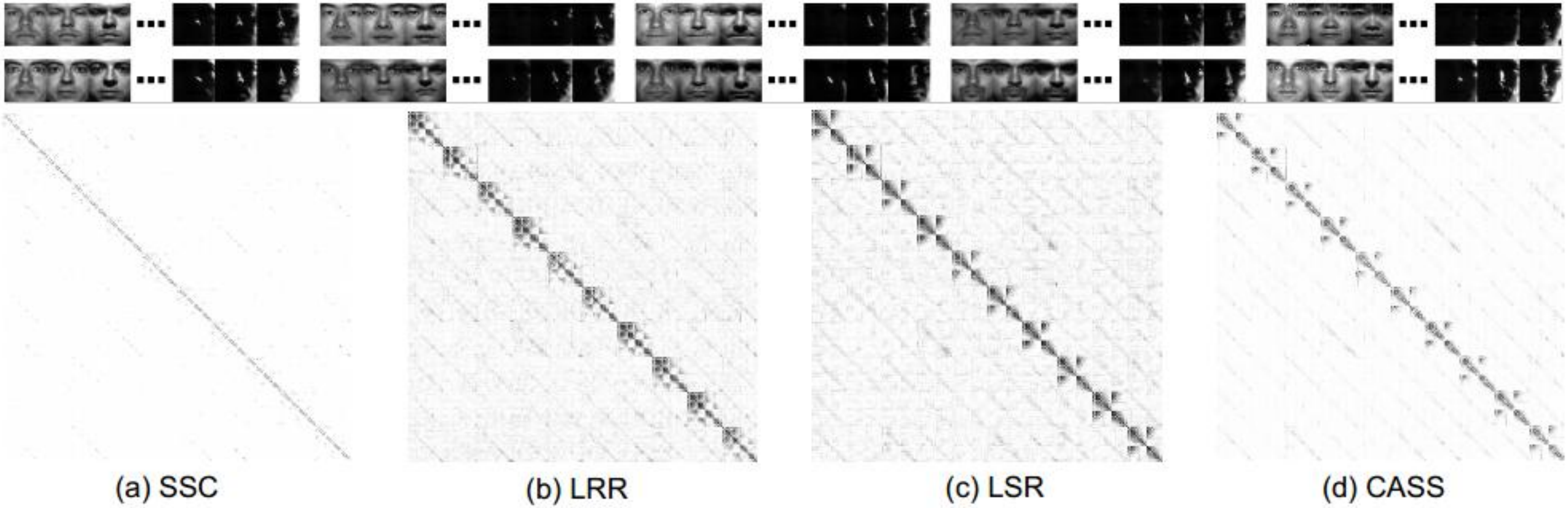
Correlation Adaptive Subspace Segmentation

- CASS also leads to block sparse solution when the data are from independent subspace.
- Grouping effect of CASS

Theorem 3 *Given a data vector* $y \in \mathbb{R}^d$, *data points* $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d \times n}$ *and parameter* $\lambda > 0$. *Let* $w^* = [w_1^*, \cdots, w_n^*]^T \in \mathbb{R}^n$ *be the optimal solution to the following problem*

$$\min_w \ \frac{1}{2}||y - Xw||_2^2 + \lambda||XDiag(w)||_*.$$

If $x_i \to x_j$, *then* $w_i^* \to w_j^*$.



Comparison of different affinity matrices

Experiments: Motion Segmentation

Table 1. The segmentation errors (%) on the Hopkins 155 database.

Comparison under the same setting					
	kNN	SSC	LRR	LSR	CASS
MAX	45.59	39.53	36.36	36.36	**32.85**
MEAN	13.44	4.02	3.23	2.50	**2.42**
STD	12.90	10.04	6.60	**5.62**	5.84

Experiments: Face Clustering

Table 2. The segmentation accuracies (%) on the Extended Yale B database.

	kNN	SSC	LRR	LSR	CASS
5 subjects	56.88	80.31	86.56	92.19	**94.03**
8 subjects	52.34	62.90	78.91	80.66	**91.41**
10 subjects	50.94	52.19	65.00	73.59	**81.88**

Thanks!