Robust and Efficient Subspace Segmentation via Least Squares Regression

Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan

Subspace Segmentation Problem

Given a set of data vectors $X = [X_1, \cdots, X_k] = [x_1, \cdots, x_n] \in \mathbb{R}^{d \times n}$ drawn from a union of k subspaces $\{\mathcal{S}_i\}_{i=1}^k$. Let X_i ba a collection of n_i data vectors drawn from the subspace $\mathcal{S}_i$, $n = \sum_{i=1}^k n_i$. The task is to segment the data according to the underlying subspaces they are drawn from.

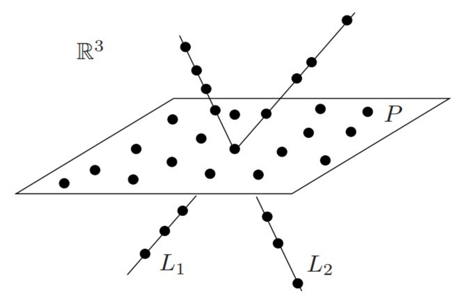

Figure 1: *A set of sample points in $\mathbb{R}^3$ drawn from a union of three subspaces: two lines and a plane.*

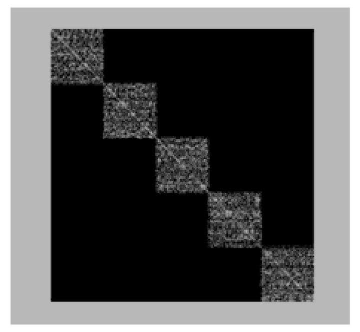

Figure 2: *The ideal affinity matrix is* ***block diagonal***.

Related Works

- Spectral Clustering (SC) is used as the framework for subspace segmentation.
- The main challenge by using SC is to define a "good" affinity matrix (or graph) $Z \in \mathbb{R}^{n \times n}$. Each entry Z_{ij} measures the similarity between x_i and x_j.
- Ideally, the affinity matrix should be **block diagonal**, the between-cluster affinities are all zeros (See **Figure 2**).

SSC (Sparse Subspace Clustering) (Elhamifar and Vidal, CVPR 2009):

$$\min ||Z||_0 \text{ s.t. } X = XZ, \ \operatorname{diag}(Z) = 0. \quad (1)$$

$$\min ||Z||_1 \text{ s.t. } X = XZ, \ \operatorname{diag}(Z) = 0. \quad (2)$$

If the subspaces are **independent**[1], the solution Z^* to (2) is **block diagonal**.

LRR (Low-Rank Representation) (Liu et al., ICML 2010, TPAMI 2012):

$$\min \operatorname{rank}(Z) \text{ s.t. } X = XZ. \quad (3)$$

$$\min ||Z||_* \text{ s.t. } X = XZ. \quad (4)$$

If the subspaces are **independent**, the solution Z^* to (4) is **block diagonal**.

MSR (Multi-Subspace Representation) (Luo et al., ECML PKDD 2011):

$$\min ||Z||_* + \delta ||Z||_1 \text{ s.t. } X = XZ, \operatorname{diag}(Z) = 0. \quad (5)$$

If the subspaces are **independent**, the solution Z^* to (5) is **block diagonal**.

SSQP (Subspace Segmentation via Quadratic Programming) (Wang et al., AAAI 2011):

$$\min ||XZ - X||_F^2 + \lambda ||Z^T Z||_1 \text{ s.t. } Z \geq 0, \operatorname{diag}(Z) = 0. \quad (6)$$

If the subspaces are **orthogonal**, the solution Z^* to (6) is **block diagonal**.

[1] A collection of k subspaces $\{\mathcal{S}_i\}_{i=1}^k$ are independent if and only if $\sum_{i=1}^k \mathcal{S}_i = \oplus_{i=1}^k \mathcal{S}_i$.

[2] The data sampling is sufficient which makes the problem (7) have a nontrivial solution.

Theoretical Analysis

Consider a general model as follow:

$$\min f(Z) \text{ s.t. } Z \in \Omega = \{Z | X = XZ\}. \quad (7)$$

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

Theorem 1 *Assume the data sampling is sufficient* [2]*, and the subspaces are* ***independent***. *If f satisfies the EBD conditions (1)(2), the optimal solution(s) Z^* to problem (7) 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:

$$\min f(Y) \text{ s.t. } X_i = X_i Y. \quad (8)$$

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

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

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

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

must be ***block diagonal***, *where $p > 0$, and $\lambda > 0$.*

Some Remarks:

- EBD (1) is the basic requirement for subspace segmentation; EBD (2) enforces the solution to be block diagonal; EBD (3) shows the representation coefficient in each block.
- For SSC, the solution may be too sparse if the data are highly correlated.
- For LRR by (3), $rank(Z)$ does not satisfies the EBD (2).

LSR (Least Squares Regression)

LSR without noise: $\min ||Z||_F \text{ s.t. } X = XZ, \operatorname{diag}(Z) = 0.$ (10)

LSR1 with noise: $\min ||X - XZ||_F^2 + \lambda ||Z||_F^2 \text{ s.t. } \operatorname{diag}(Z) = 0.$ (11)

LSR2 with noise: $\min ||X - XZ||_F^2 + \lambda ||Z||_F^2.$ (12)

The Grouping Effect: LSR exhibits the grouping effect that the coefficients of a group of correlated data are approximately equal:

Theorem 3 *Given a data vector $y \in \mathbb{R}^d$, data points $X \in \mathbb{R}^{d \times n}$ and a parameter λ. Assume each data point of X are normalized. Let z^* be the optimal solution to the following LSR (in vector form) problem:*

$$\min ||y - Xz||_2^2 + \lambda ||z||_2^2. \quad (13)$$

We have

$$\frac{||z_i^* - z_j^*||_2}{||y||_2} \leq \frac{1}{\lambda} \sqrt{2(1-r)}, \quad (14)$$

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

Some Remarks:

- The grouping effect of LSR shows that highly correlated data tends to be grouped in a same cluster.
- SSC and LRR by (3) does not has the grouping effect.
- LRR by (4) has the grouping effect.

Experimental Verification

Table 3: *Comparison of the segmentation errors (%) and running time (s) on the Hopkins 155 Database*

	SSC	LRR	LSR1	LSR2
Max	39.53	36.36	36.36	36.36
Mean	4.02	3.23	**2.50**	2.84
Median	0.90	0.50	**0.31**	0.34
STD	10.04	6.06	**5.62**	6.16
Time	149.70	129.30	24.33	**21.35**

Table 4: *Comparison of the segmentation accuracies (%) and running time (s) on the Extended Yale Database B*

		SSC	LRR	LSR1	LSR2
Acc.	5	76.88	81.88	88.13	**91.56**
	10	47.81	65.00	70.16	**72.34**
Running time	5	0.657	0.602	0.018	**0.009**
	10	4.760	2.261	0.101	**0.045**

Conclusions

- The proposed EBD conditions summaries some existing methods.
- The grouping effect is important for subspace segmentation, especially for correlated data. The proposed LSR is effective and efficient.
- LSR is simpler and better. The previous models, SSC, LRR, MSR and SSQP, are unnecessary sophistication.