"Sparce Subspace Clustering Algorithm, Theory and Applications"

Report by:

Mohsen Nabian

- Subspace clustering problem:

given points $\{y_1, y_2, \ldots y_N\}$ in $\mathbb{R}^n$ lying in subspaces

$S_1 \cup \ldots \cup S_L$, find:

- Basis for each subspace
- clustering of the data

Challenges:

- do not know subspace bases
- do not know membership of points
- ~~The~~ data points ~~...~~ may be Corrupted by noise, missing enteries, outliers, ...

- Self expressiveness property of the data ~~...~~ :

"each data point in a union of subspaces can be efficiently reconstructed by a combination of other points in the dataset."

or each data point $y_i \in \cup_{l=1}^{n} S_l$ can be written as:

all data points

$$\underset{n\times 1}{[y_i]} = \underset{n\times N}{[Y]}\underset{N\times 1}{[c_i]}, \quad c_{ii} = 0 \qquad c_i \triangleq [c_{i1}, c_{i2}, \ldots c_{iN}]^T$$

one data point

Y

since for $[Y]_{n\times N}$, $N>n$, $[Y]_{n\times N}$ is a fat matrix and therefore $y_i = Y c_i$ has many solutions. however:

"there exist a sparce solution c_i, whoes nonzero enteries correspond to data points from the same subspace as y_i. we refer to such a solution as a subspace-sparse representation."

Thus, from the system of eq $\overset{\checkmark}{y_i} = \overset{\checkmark ?}{Y} c_i$, with infinitely many solutions, one can restrict set of solution by

$$\min \overset{l_q\text{-norm}}{\|c_i\|_q} \quad \text{s.t.} \quad y_i = Y c_i \,, \quad c_{ii} = 0$$

$q \downarrow$ sparcity of $c_i \uparrow$

extreme case $q=0 \Rightarrow$ NP-hard

So we use tightest convex relaxation of l_0-norm:

$$\min \|c_i\|_1 \quad \text{s.t.} \quad y_i = Y c_i \quad c_{ii} = 0 \qquad (\mathrm{I})$$

writing (I) for all data points, $i=1,\dots,N$:

$$\min \|C\|_1 \quad \text{s.t.} \quad Y = YC \, , \; \text{diag}(C) = 0 \qquad (\text{II})$$

Clustering using sparce coefficients:

after solving II, we obtained sparce $\underline{C}$.
now we have to ~~classify~~ cluster the data points.
So we build a weighted graph $G = (V, \mathcal{E}, W)$
where V denots the set of N nodes (N data points),
$\mathcal{E} \subseteq V \times V$, set of edges between nodes and
$W \in \mathbb{R}^{N\times N}$, a symmetric similarity matrix representing
the weights of the edges. W is defined as:

$$W = |C| + |C^T| \quad \text{or} \quad w_{ij} = |c_{ij}| + |c_{ji}|$$

By forming graph G, we use spectral clustering
to cluster data points.

Sparce Subspace clustering (SSC) (Algorithm I)

input: A set of points $\{y_i\}_{i=1}^N$ lying in a union of n linear subspace $\{S_i\}_{i=1}^n$.

1) Solve the sparse optimization II $\Rightarrow C$ ✓

2) Normalize C by columns

3) Form similarity graph with N nodes.

$$W = |C| + |C^T|$$

4) Apply spectral clustering

output: segmentation of the data: $Y_1, Y_2, \ldots, Y_n$

Dealing with Noise and and sparse Outlying Entries:

$$\underset{\text{ith data points}}{y_i} = \underset{\text{error free point}}{y_i^0} + \underset{\text{sparce outlying entery}}{e_i^0} + \underset{\text{noise}}{z_i^0} \qquad \|e_i^0\|_0 \leq k$$

self expressive ness property: $y_i^0 = \sum_{j \neq i} c_{ij} y_j^0$

$$\left.\begin{aligned} e_i &\triangleq e_i^0 - \sum_{j\neq i} c_{ij} e_j^0 \\ z_i &\triangleq z_i^0 - \sum_{j\neq i} c_{ij} z_j^0 \end{aligned}\right\} \Rightarrow y_i = \sum_{j\neq i} c_{ij} y_j + e_i + z_i$$

for all data points:

$$Y = YC + E + Z \quad diag(C) = 0$$

$\Rightarrow$ So we modify (II) to:

$$\begin{cases} \min \|C\|_1 + \lambda_e \|E\|_1 + \frac{\lambda_z}{2} \|Z\|_F^2 \\ s.t. \quad Y = YC + E + Z \end{cases} \qquad (III)$$

Suggestions: $\lambda_z = \frac{\alpha_z}{\mu_z}$ $\quad \lambda_e = \frac{\alpha_e}{\mu_e}$ $\quad \alpha_z, \alpha_e > 1$

$$\mu_z \triangleq \min_i \max_{j \neq i} |y_i^T y_j| \qquad \mu_e \triangleq \min_i \max_{j \neq i} \|y_j\|_1$$

So with this l_1 minimization in III we are able to find $((C, E, Z))$ and therefore clustering $\underline{C}$.

✓

Missing Enteries

Consider a collection $\{y_i\}_{i=1}^N \in \mathbb{R}^D$

let $J_i \subset \{1, \dots, D\}$ denote indices of known entery of y_i.

let $J \triangleq \bigcap_{i=1}^N J_i$

if $|J|$ is not small relative to D, then (Next Page) we can implement SSC algorithm on

$\bar{Y} \in \mathbb{R}^{|J| \times N}$, where $\bar{Y}$ is reduced dim data points where

from Y where all enteries are known.

However if $|J|$ is very small or $\approx$ zero

this algorithm does not hold.

Theoritical Analysis:

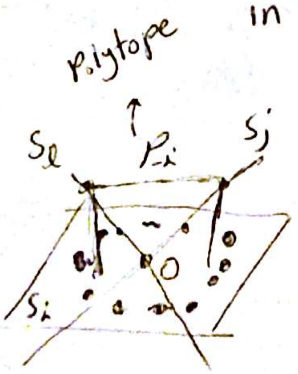
Δ SSC succeeds when:

- l_1 selects points from the correct subspace or (no false discovery)

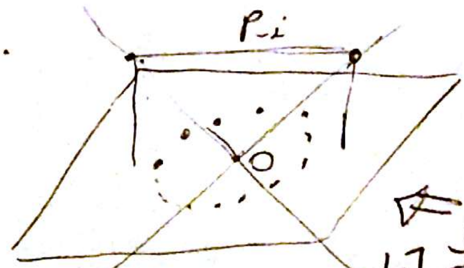
SSC has zero false discovery for any $y \in S_i$

if $\max_{j \neq i} C_s(\theta_{ij}) < \max_{\mathrm{rank}(Y_i') = d_i} \sigma_{d_i}(Y_i') / \sqrt{d_i}$

in other words: No need to many points. Need a few but well distributed



$\Leftarrow$ L1 Succeed



$\Leftarrow$ L1 fails

So the success of l_1-minimization for subspace-sparse recovery depends on the principal angle between subspaces and the distribution of the data in each subspace.

Experimental Results:

~~Syn~~ (1) synthetic data:

Create ~~consider~~ 3 disjoint subspaces $\{S_i\}_{i=1}^{3}$ Same dimension d. such that smallest principal angle is $\theta = \theta_{12} = \theta_{23}$

also in each subspace, we put N_g random data points.

we endup with:

$$\left| \begin{array}{l} \text{increasing } \theta \\ \\ \text{increasing } N_g \end{array} \right. \Rightarrow \text{reduces} \left\{ \begin{array}{l} \text{Subspace sparse recovery error} \\ \\ \text{Subspace clustering error} \end{array} \right.$$

imagin that for y_i, follwing sparse representation is computed: $y_i \to c_i^T = [c_{i1}^T \; c_{i2}^T \; c_{i3}^T]$

⇨ Subspace-sparse recovery error is defined as follows:

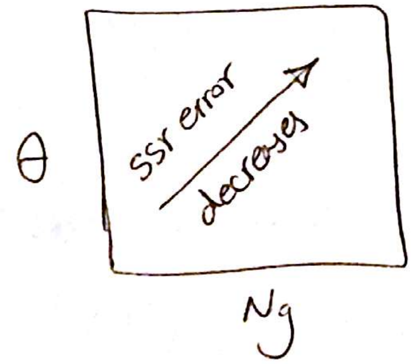
$$\text{ssr error} = \frac{1}{3N_g} \sum_{i=1}^{3N_g} \left(1 - \frac{\|c_{i\,k_i}\|_1}{\|c_i\|_1}\right) \in [0,1]$$

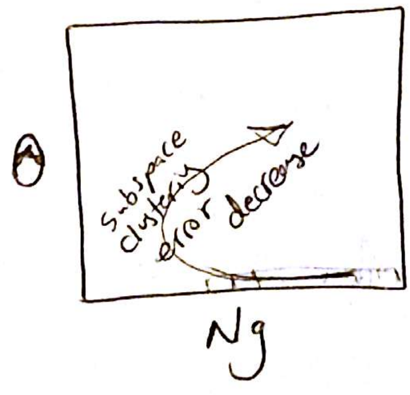
$$\text{Subspace clustering error} = \frac{\#\text{ of misclassified points}}{\text{total } \#\text{ of points}}$$

in experiment, dimension of ambient space = $D = 50$

$\theta \in [6, 60]$ $\quad N_g \in [d+1, 32d]$

result of $d = 4$ is provided.





Experiment with real data

- Tow problens → segmenting multiple motions in video.
  → clustering images of human faces.

- Compapre SCC with:

  LSA, SCC, LRR, LRSC

  state-of-the-art subspace clustering algorithms.

- for motion segmentation: → Hopkins 155 dataset which consist of 155 video sequences of 2 or 3 motions corresponding to 2 or 3 low dimensional subspaces in each video.

- for Extended Yale B data set, 38 humans

angles between subspaces: θ
- Hpkin → $\theta \leq 10$
- Yale → $10 \leq \theta \leq 20$

k nearest neighbors
- Hpkin → almost all data points k nearest neighbors belong to same cluster
- Yale → not good, k↑ worse

Motion Segmentation:

N feature points through frames $f = 1, \ldots, F$ } $y_i \triangleq [x_{1i}^T, x_{2i}^T, \ldots x_{Fi}^T] \in \mathbb{R}^{2F}$

120 videos $\xrightarrow{\text{frames}}$ 2 motions   N = 266   30 frames

35 " $\Rightarrow$ 3 "   N = 398   29 "

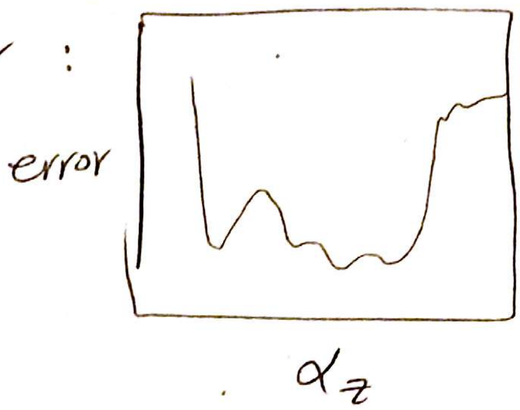
result

Algo	LSA	SCC	LRR	LRR-H	LRSC	SCC
2 Motion error						1.52
3 Motion error						4.40
together (All) error						2.18

- Table T shows Scc is the most successful in identifying subspaces in comparison with others.
- dividing into 2 motions & 3 motions, improved accuracy for 2 motion.

- Same observations happens when we use 4n dimensional data points obtained by applying PCA (Table 2)

Fig 11 also showed the effect of regularization parameter:



$$\lambda_z = \frac{\alpha_z}{\mu_z}$$

Face Clustering

Table 3: after applying RPCA separately to the data points in each subjects $\Rightarrow$ LRSC is the best
SSC is 2nd best
(best: lowest clustering error)

Table 4: apply RPCA simultaneously to all data in each trial $\Rightarrow$ SSC has lowest clustering error in all trials
2, 3, 5, 8 & 10 subjects.

Table 5 : No preprocessing => SSC has Lowest clustering error in all trials: 2, 3, 5, 8 & 10 subjects.

Computational Time :

average Computational Time ploted on Yale-B dataset as a function of the number of subjects :