Microsoft

CVPR

微软亚洲研究院创研论坛

CVPR 2020 论文分享会

Stochastic Sparse Subspace Clustering

When SSC meets Dropout …

Ying Chen (谌瑛) [1], Chun-Guang Li (李春光) [1], Chong You (由翀) [2]

[1] PRIS Lab., SCIE, Beijing University of Posts and Telecommunications (北京邮电大学)
[2] EECS, University of California, Berkeley

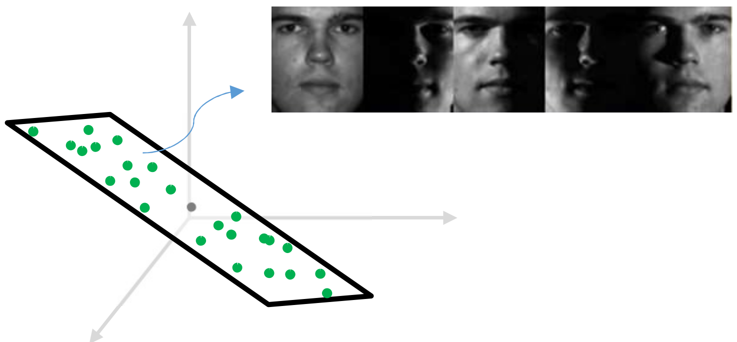
Outline

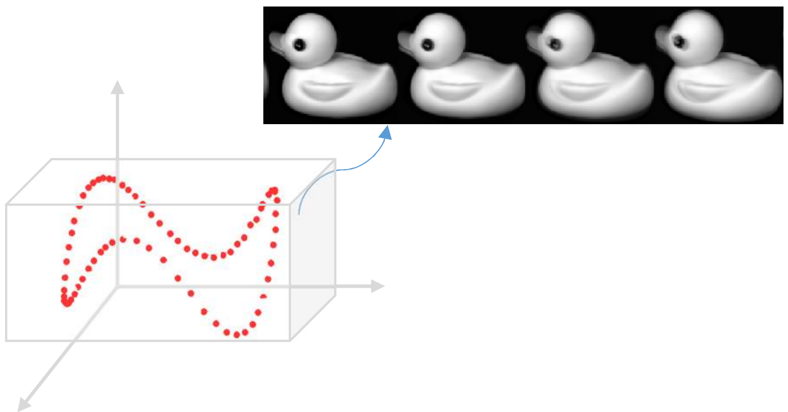
- Introduction to Subspace Clustering
- Related Work and Motivation
- Our Proposal
- Experiments
- Summary

Structure in High-Dimensional Data

- High dimensional data belonging to a class (or category) are usually controlled by a few hidden factors

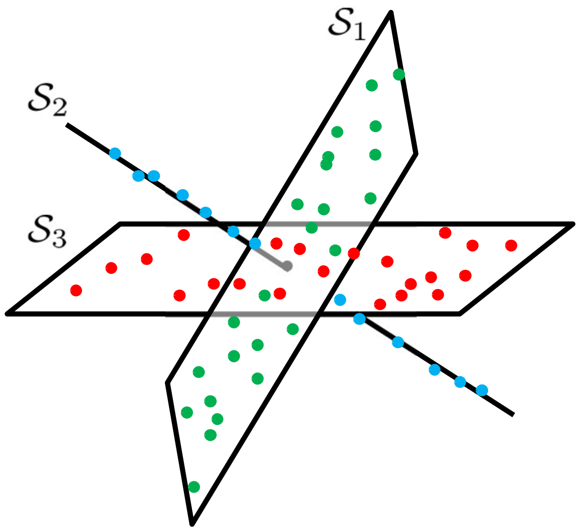
  - Well approximated by a linear subspace or a nonlinear submanifold, e.g.:





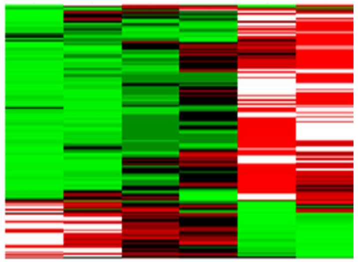
Structures in High-Dimensional Data

- High dimensional data with multiple classes (or categories) usually lie in multiple low-dimensional structures
  - ➢ We approximate the low-dimensional structures by a union of subspaces

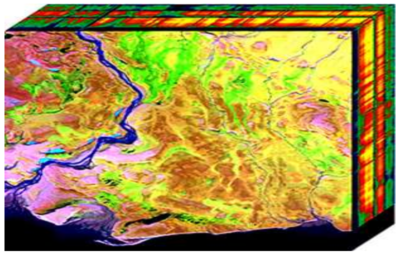


e.g.:

☑ Feature point trajectories of rigid object in video



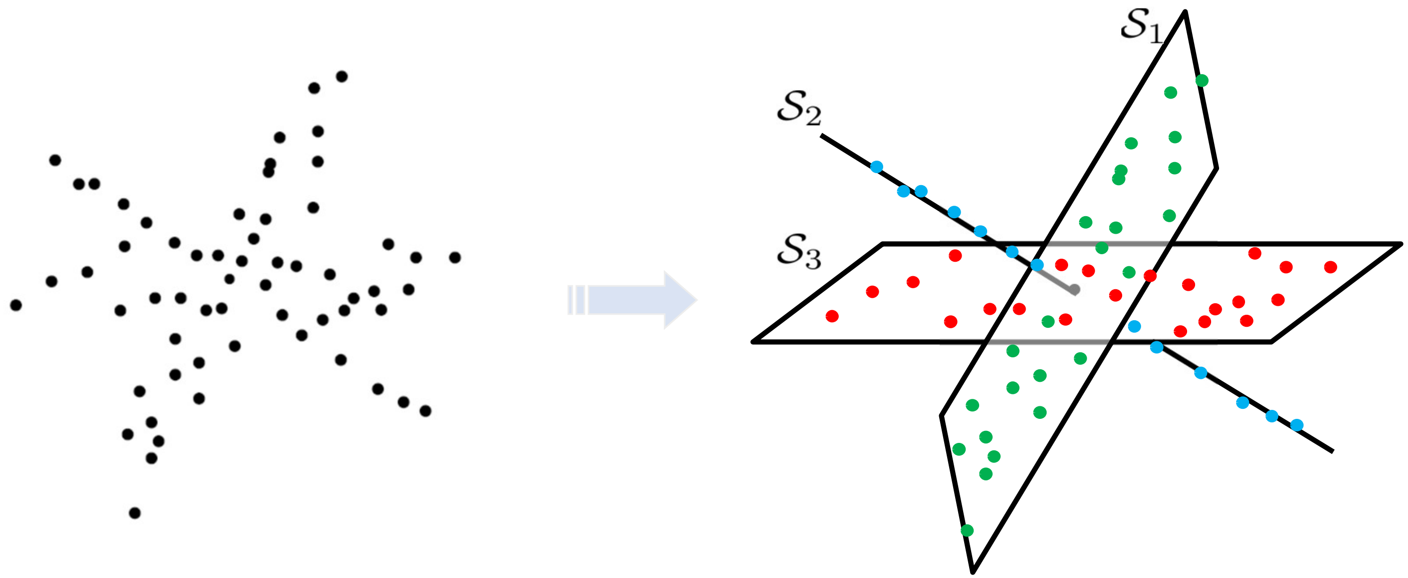
☑ Gene expression array in bioinformatics

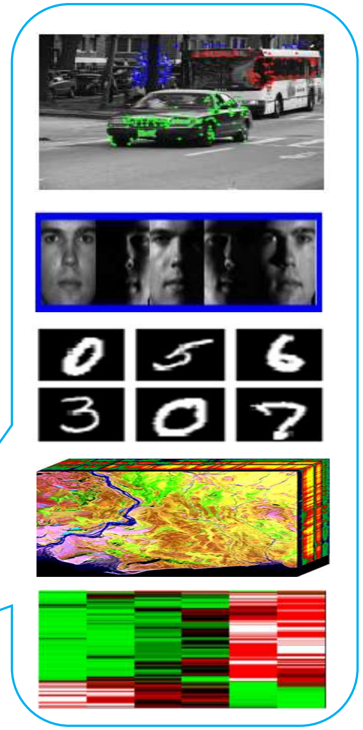


☑ Hyperspectral image in remote sensing

Subspace Clustering

- Given a set of data points lying (approximately) in a **union of subspaces**, to segment the data points into each subspace



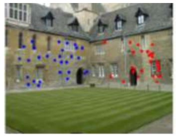


- A subspace corresponds to a pattern (or cluster) in data set
  - e.g. a moving object, an individual, a digit, a type of area, a cancer subtype

Segmenting Data into Subspaces → Subspace Clustering

- Rigid Object Segmentation in Video

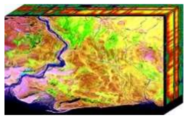
- Planar Area Detection in 3D



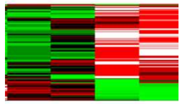
- Face Image Clustering

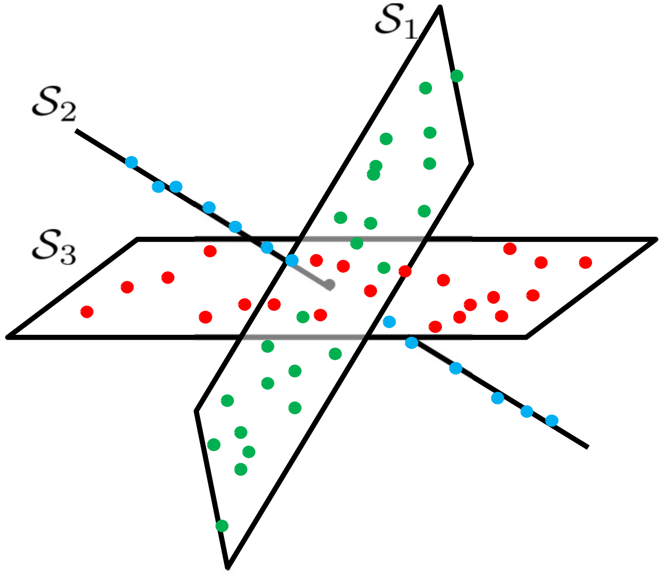
- Handwriting Digit Image Clustering

- Hyperspectral Image Segmentation



- Cancer Subtype Discovery





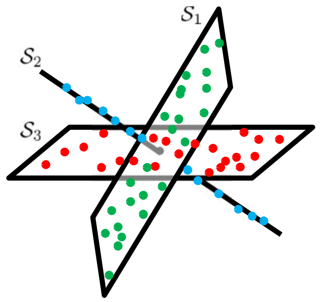
Union of Subspaces

Outline

- Introduction to Subspace Clustering
- Related Work and Motivation
- Our Proposal
- Experiments
- Summary

Subspace Clustering: Existing Methods

- Iterative Methods
  - k-plane (Bradley & Manga.: J. Global Opt. ' 00), q-flats, ···

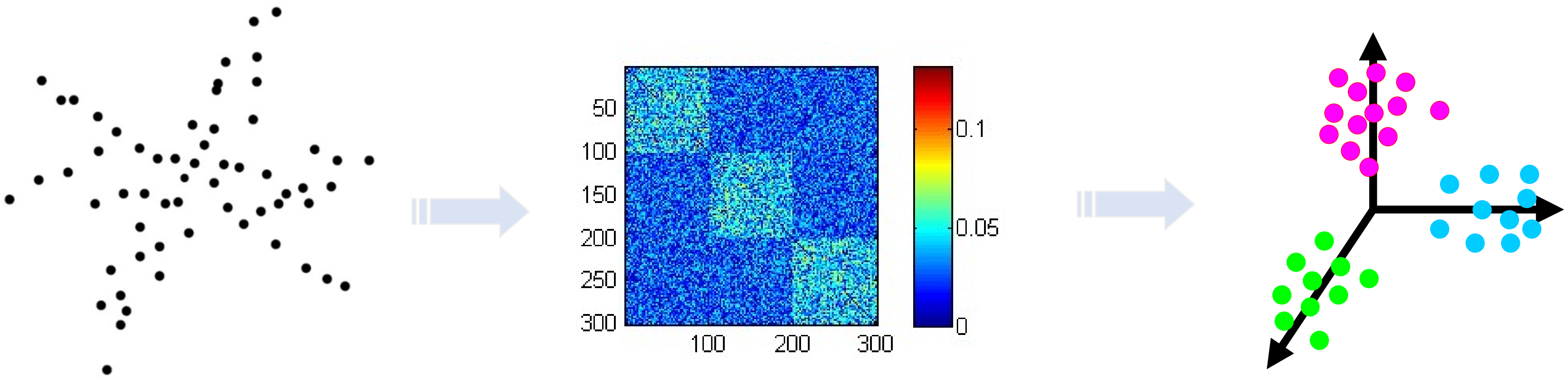


- Statistical Methods
  - Matrix Factorization, MPPCA, ···
- Algebraic Geometry based Methods
  - Generalized PCA (GPCA), ···
- Spectral Clustering based Methods
  - SSC (Elhamifar & Vidal: CVPR09), LRR (Liu et al. ICML10), LSR (Lu et al. ECCV12)
  - Kernel SSC/Latent SSC(Vishal et al. ICCV13), Latent LRR(Liu et al. ICCV11), VR3(Peng et al. CVPR17), DSCNet (Ji et al.:NIPS17), DASC(Zhou et al.: CVPR18), S^2ConvSCN(Zhang et al. CVPR19)
  - Correntropy (He et al. TNNLS15), GMM (Li et al.:CVPR15), Weighted Error Entropy (Li et al.: CVPR19)
  - Scalable subspace clustering: SSSC (Peng et al. CVPR13), OLRSC(Shen et al. ICML16), EnSC (You et al. CVPR16), SSCOMP (You et al. CVPR16), ESC (You et al. ECCV18), SR-SSC (Abdolali et al. SP19)

[1] P. S. Bradley & Mangasarian. k-plane clustering. Journal of Global Optimization, 16(1):23–32, 2000.

Spectral Clustering based Methods

- Step 1: Build a Data Affinity Matrix
- Step 2: Apply Spectral Clustering



- Why spectral clustering?
  - ✓ Spectral Graph Theory

[1] R. Vidal. Subspace clustering. IEEE Signal Processing Magazine, 28(3):52–68, March 2011.
[2] C.-G. Li, C. You, & R. Vidal, "Structured Sparse Subspace Clustering: A Joint Affinity Learning and Subspace Clustering Framework", IEEE Transactions on Image Processing, 2017.

Self-Expression Model

- Step 1: Build a Data Affinity Matrix
- Step 2: Apply Spectral Clustering

- Self-Expression Model
  - Express a data point $\boldsymbol{x}_j$ as a linear combination of other data points

$$\boldsymbol{x}_j = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_{j-1}\boldsymbol{x}_{j-1} + 0 \cdot \boldsymbol{x}_j + c_{j+1}\boldsymbol{x}_{j+1} + \cdots + c_N\boldsymbol{x}_N$$

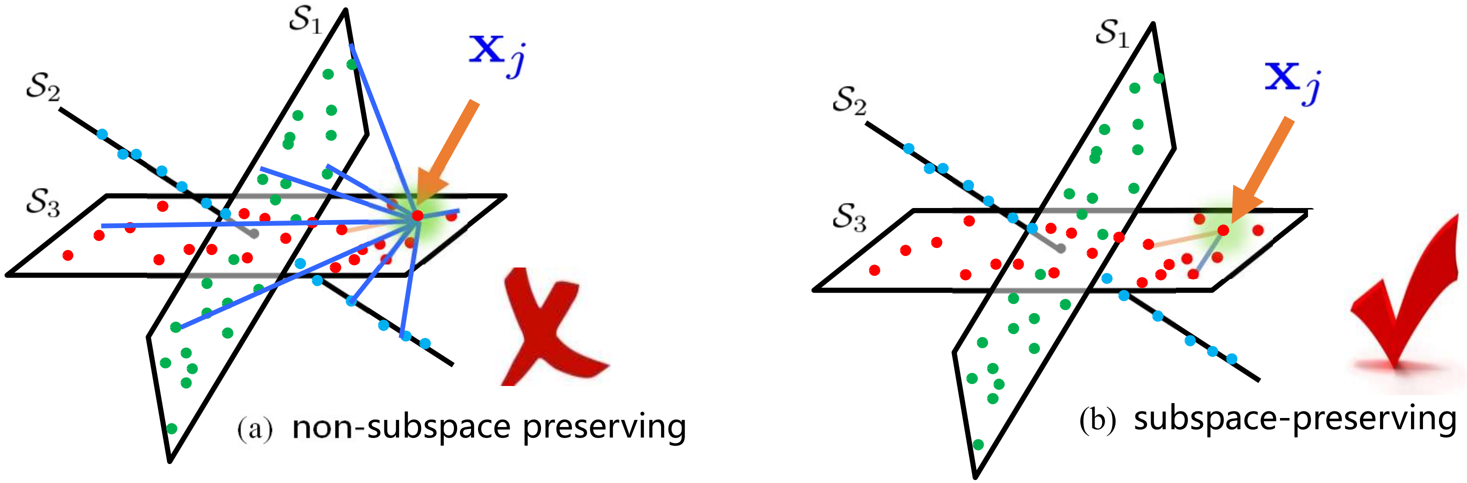
[1] E. Elhamifar & R. Vidal: "Sparse subspace clustering", CVPR 2009.
[2] E. Elhamifar & R. Vidal: "Sparse subspace clustering: Algorithm, theory, and applications", IEEE TPAMI 2013.

Subspace Preserving Property

- Self-Expression Model
  - For data points in a linear subspace, we have:

$$\boldsymbol{x}_j = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_{j-1}\boldsymbol{x}_{j-1} + 0 \cdot \boldsymbol{x}_j + c_{j+1}\boldsymbol{x}_{j+1} + \cdots + c_N\boldsymbol{x}_N$$

  - **Subspace-Preserving Property:** nonzero coefficients ONLY correspond to data points in the same subspace as $\boldsymbol{x}_j$



(a) non-subspace preserving

(b) subspace-preserving

Subspace Preserving Property

- Self-Expression Model
  - For data points in a linear subspace, we have:

a linear combination

$$x_j = c_1 x_1 + c_2 x_2 + \cdots + c_{j-1} x_{j-1} + 0 \cdot x_j + c_{j+1} x_{j+1} + \cdots + c_N x_N$$

  - To find the coefficients with **subspace-preserving property**, a proper **regularization** is needed, e.g.

✓ $\|c\|_1$ → Sparse Subspace Clustering (SSC)

✓ $\|c\|_2^2$ → Least Square Regression (LSR)

✓ $\|C\|_*$ → Low-Rank Representation (LRR/LRSC)

Theoretical guarantees of using ℓ_1 norm: Independent, disjoint, **affine**, even with outliers, noisy, missing entries, and etc.

[1] E. Elhamifar & R. Vidal: CVPR'09. [2] Sol. & Candes: Annual Stat.'12. [3] E. Elhamifar & R. Vidal: TPAMI '13. [4] Y.-X. Wang & H. Xu: ICML' 13; JMLR' 16. [5] M. Tsakiris et al. ICML'18. [6] C.-G. Li, C. You, & R. Vidal, SPARS'15 / IEEE J. STSP '18.

Spectral Clustering based Methods: Regularization on C

$$\min_{C,E} \|C\|_\kappa + \lambda\|E\|_\ell \quad \text{s.t.} \quad X = XC + E, \quad \text{diag}(C) = 0$$

- Spectral Clustering based Methods
  - ℓ_1 in SSC (Elhamifar & Vidal: CVPR09), $\|C\|_*$ in LRR (Liu et al. ICML10), ℓ_2 in LSR (Lu et al. ECCV12)
  - Kernel SSC/Latent SSC(Vishal et al. ICCV13), Latent LRR(Liu et al. ICCV11), VR3(Peng et al. CVPR17), DSCNet (Ji et al.:NIPS17), DASC(Zhou et al.: CVPR18), S^2ConvSCN(Zhang et al. CVPR19)
  - Correntropy (He et al. TNNLS15), GMM (Li et al.:CVPR15), Weighted Error Entropy (Li et al.: CVPR19)
  - Scalable subspace clustering: SSSC (Peng et al. CVPR13), EnSC (You et al. CVPR16), SSCOMP (You et al. CVPR16), OLRSC (Shen et al. ICML16) , ESC (You et al. ECCV18), SR-SSC (Abdolali et al. SP19)

Spectral Clustering based Methods: Feature Space

$$\min_{C,E,\Theta} \|C\|_\kappa + \lambda\|E\|_\ell \quad \text{s.t.} \quad \phi_\Theta(X) = \phi_\Theta(X)C + E, \quad \text{diag}(C) = 0$$

- Spectral Clustering based Methods
  - ℓ_1 in SSC (Elhamifar & Vidal: CVPR09), $\|C\|_*$ in LRR (Liu et al. ICML10), ℓ_2 in LSR (Lu et al. ECCV12)
  - Kernel SSC/Latent SSC(Vishal et al. ICCV13), Latent LRR(Liu et al. ICCV11), VR3(Peng et al. CVPR17), DSCNet (Ji et al.:NIPS17), DASC(Zhou et al.: CVPR18), S^2ConvSCN(Zhang et al. CVPR19)
  - Correntropy (He et al. TNNLS15), GMM (Li et al.:CVPR15), Weighted Error Entropy (Li et al.: CVPR19)
  - Scalable subspace clustering: SSSC (Peng et al. CVPR13), EnSC (You et al. CVPR16), SSCOMP (You et al. CVPR16), OLRSC (Shen et al. ICML16), ESC (You et al. ECCV18), SR-SSC (Abdolali et al. SP19)

Spectral Clustering based Methods: Error Modeling

$$\min_{C,E} \|C\|_\kappa + \lambda \|E\|_\ell \quad \text{s.t.} \quad X = XC + E, \quad \text{diag}(C) = 0$$

- Spectral Clustering based Methods
  - ℓ_1 in SSC (Elhamifar & Vidal: CVPR09), $||C||_*$ in LRR (Liu et al. ICML10), ℓ_2 in LSR (Lu et al. ECCV12)
  - Kernel SSC/Latent SSC(Vishal et al. ICCV13), Latent LRR(Liu et al. ICCV11), VR3(Peng et al. CVPR17), DSCNet (Ji et al.:NIPS17), DASC(Zhou et al.: CVPR18), S^2ConvSCN(Zhang et al. CVPR19)
  - Correntropy (He et al. TNNLS15), GMM (Li et al.:CVPR15), Weighted Error Entropy (Li et al.: CVPR19)
  - Scalable subspace clustering: SSSC (Peng et al. CVPR13), EnSC (You et al. CVPR16), SSCOMP (You et al. CVPR16), OLRSC (Shen et al. ICML16), ESC (You et al. ECCV18), SR-SSC (Abdolali et al. SP19)

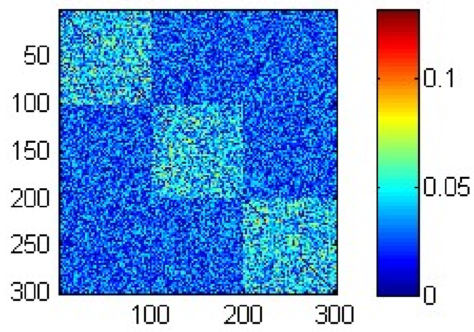
Spectral Clustering based Methods: Scalability

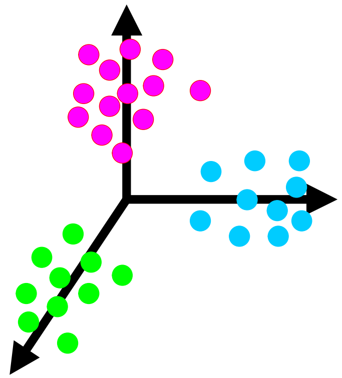
$$\min_{C,E} \|C\|_\kappa + \lambda\|E\|_\ell \quad \text{s.t.} \quad X = XC + E, \quad \text{diag}(C) = 0$$

- Spectral Clustering based Methods
  - ℓ_1 in SSC (Elhamifar & Vidal: CVPR09), $||C||_*$ in LRR (Liu et al. ICML10), ℓ_2 in LSR (Lu et al. ECCV12)
  - Kernel SSC/Latent SSC(Vishal et al. ICCV13), Latent LRR(Liu et al. ICCV11), VR3(Peng et al. CVPR17), DSCNet (Ji et al.:NIPS17), DASC(Zhou et al.: CVPR18), S^2ConvSCN(Zhang et al. CVPR19)
  - Correntropy (He et al. TNNLS15), GMM (Li et al.:CVPR15), Weighted Error Entropy (Li et al.: CVPR19)
  - Scalable subspace clustering: SSSC (Peng et al. CVPR13), EnSC (You et al. CVPR16), SSCOMP (You et al. CVPR16), OLRSC (Shen et al. ICML16), ESC (You et al. ECCV18), SR-SSC (Abdolali et al. SP19)

Spectral Clustering based Methods: Connectivity

- Step 1: Build a Data Affinity Matrix
- Step 2: Apply **Spectral Clustering**





- Why spectral clustering?
  - ✓ Spectral Graph Theory

1. Nonzero entries in affinity matrix are correct (i.e. subspace preserving)
2. Affinity graph corresponding to data points from the same subspace is well-connected.

[1] R. Vidal. Subspace clustering. IEEE Signal Processing Magazine, 28(3):52–68, March 2011.
[2] C.-G. Li, C. You, & R. Vidal, "Structured Sparse Subspace Clustering: A Joint Affinity Learning and Subspace Clustering Framework", IEEE Transactions on Image Processing, 2017.

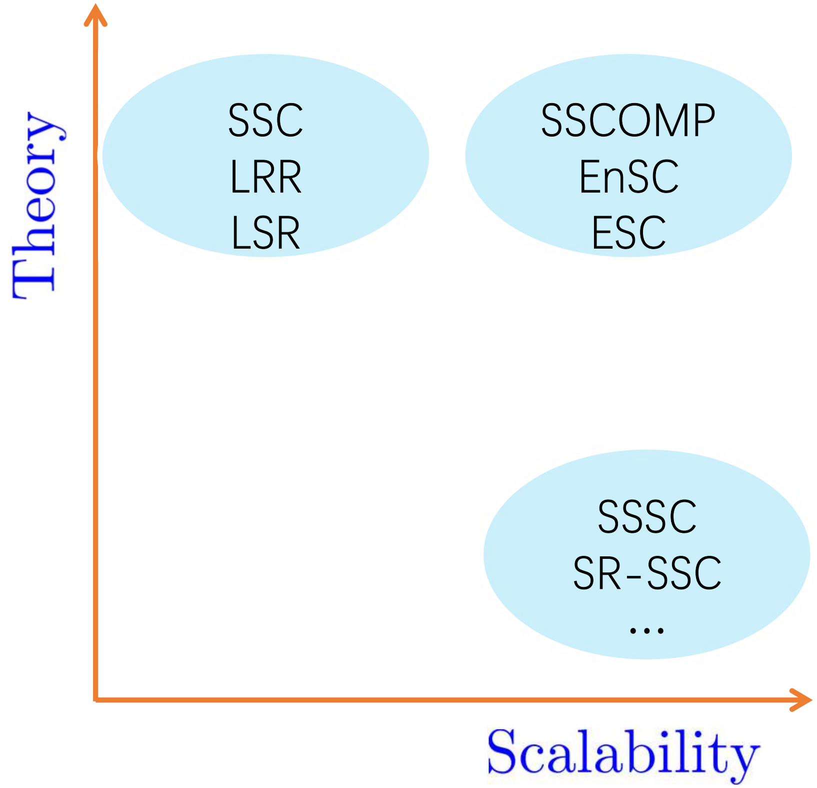
Connectivity Issue in Sparse Subspace Clustering

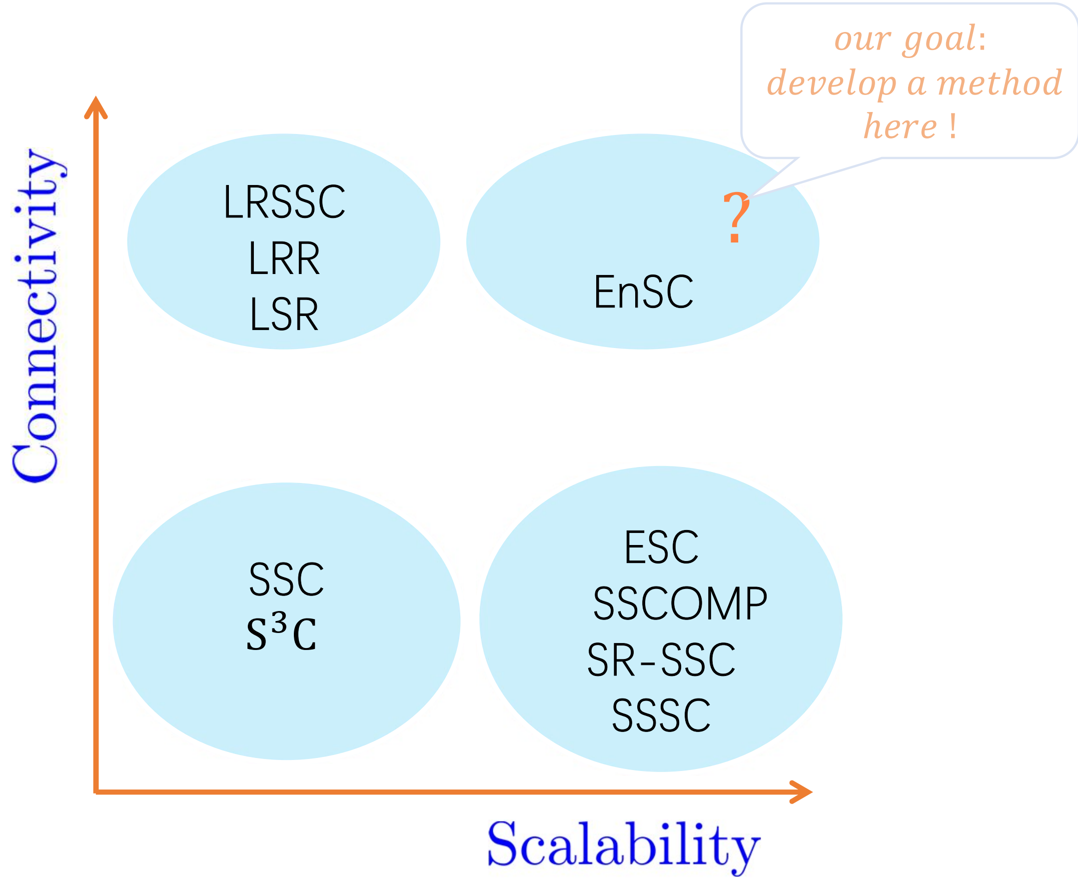
- When dimension of subspace $d > 3$, there is a connectivity issue in SSC (Nasihatkon & Hartley, CVPR'11)
  - Data points from the same subspace form disjointed components of affinity graph, leading to erroneous over-segmentation
- Sparsity based methods suffer from connectivity issue, e.g.
  - $\mathrm{S}^3\mathrm{C}$ (Li & Vidal CVPR'15); SSCOMP (You et al. CVPR'16); ESC (You et al. ECCV'18); SR-SSC (Abdolali et al. Signal Processing '19)
- Prior Solutions:
  - Adding nuclear norm to ℓ_1 norm, e.g. LRSSC (Wang et al., NIPS'13)
  - Postprocessing by merge (Wang et al.: AISTATS'16)
  - Adding ℓ_2 norm to ℓ_1 norm, e.g. EnSC (You et al. CVPR'16)

[1] B. Nasihatkon and R. Hartley. Graph connectivity in sparse subspace clustering. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2137–2144, 2011.

Scalability, Theoretical Guarantee & Connectivity

… to provide a general approach to improve the connectivity of sparsity-based subspace clustering methods



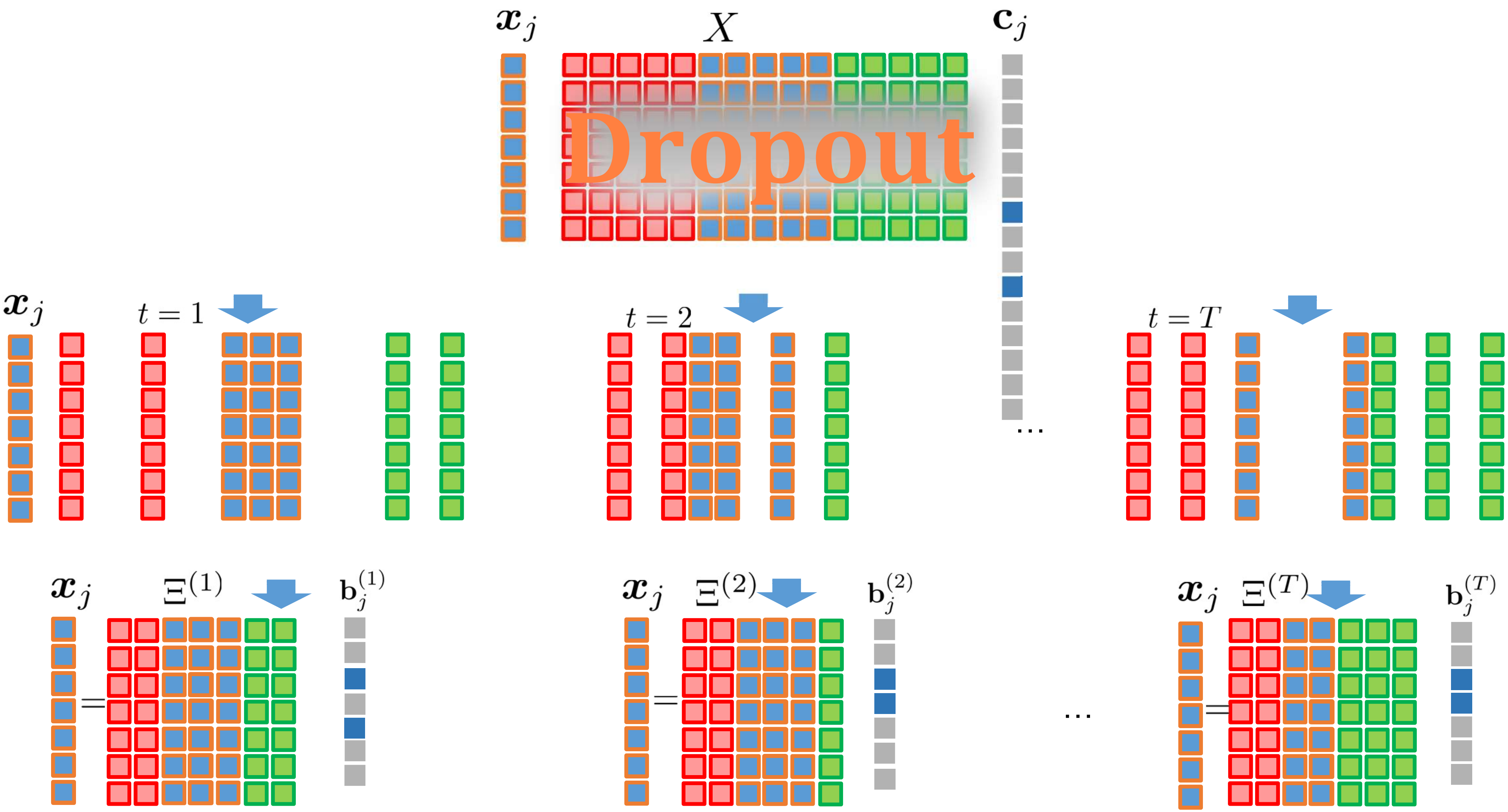


Outline

- Introduction to Subspace Clustering
- Related Work and Motivation
- Our Proposal
- Experiments
- Summary

Dropout

When SSC meets …



$\boldsymbol{x}_j$
X
$\mathbf{c}_j$
Dropout
$t=1$
$t=2$
$t=T$
$\Xi^{(1)}$
$\mathbf{b}_j^{(1)}$
$\Xi^{(2)}$
$\mathbf{b}_j^{(2)}$
$\Xi^{(T)}$
$\mathbf{b}_j^{(T)}$

Dropout in Self-Expression Model

- Define a set of Bernoulli random variables with probability distribution given by

$$\xi_i = \begin{cases} \frac{1}{1-\delta} & \text{with probability} \quad 1-\delta, \\ 0 & \text{with probability} \quad \delta. \end{cases}$$

- Introduce dropout in self-expression model

$$\min_{\boldsymbol{c}_j} \|\boldsymbol{x}_j - X\boldsymbol{c}_j\|_2^2 \quad \text{s.t.} \quad c_{jj} = 0. \Longrightarrow \min_{\boldsymbol{c}_j} \|\boldsymbol{x}_j - \sum_i \xi_i c_{ij} \boldsymbol{x}_i\|_2^2 \quad \text{s.t.} \quad c_{jj} = 0.$$

Theorem 1: Let $\{\xi_i\}_{i=1}^N$ be i.i.d. Bernoulli random variables, we have that:

$$\begin{aligned} \mathbb{E}_\xi \|\boldsymbol{x}_j - \sum_i \xi_i c_{ij} \boldsymbol{x}_i\|_2^2 &= \|\boldsymbol{x}_j - \sum_i c_{ij} \boldsymbol{x}_i\|_2^2 + \frac{\delta}{1-\delta} \sum_i \|\boldsymbol{x}_i\|_2^2 c_{ij}^2 \\ &= \|\boldsymbol{x}_j - \sum_i c_{ij} \boldsymbol{x}_i\|_2^2 + \frac{\delta}{1-\delta} \|\boldsymbol{c}_j\|_2^2 \ \text{ if } \|\boldsymbol{x}_i\|_2 = 1 \end{aligned}$$

Stochastic Sparse Subspace Clustering

- SSCOMP (You et al. CVPR16)

$$\min_{\boldsymbol{c}_j} \|\boldsymbol{x}_j - X\boldsymbol{c}_j\|_2^2, \quad \text{s.t.} \quad \|\boldsymbol{c}_j\|_0 \leq s, \quad c_{jj} = 0,$$

where $\|\cdot\|_0$ is the ℓ_0 pseudo-norm and s is a parameter that controls the sparsity

- Dropout meets SSCOMP:

$$\mathbb{E}_\xi \|\boldsymbol{x}_j - \sum_i \xi_i c_{ij} \boldsymbol{x}_i\|_2^2 \Rightarrow \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{x}_j - \sum_i \xi_i^{(t)} c_{ij} \boldsymbol{x}_i\|_2^2$$ (sample mean)

$$\min_{\boldsymbol{c}_j} \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{x}_j - \sum_i \xi_i^{(t)} c_{ij} \boldsymbol{x}_i\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{c}_j\|_0 \leq s, \quad c_{jj} = 0,$$

→ **Stochastic Sparse Subspace Clustering**

Consensus Optimization Problem

- Consensus optimization problem

$$\min_{\boldsymbol{c}_j} \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{x}_j - \sum_i \xi_i^{(t)} c_{ij}\boldsymbol{x}_i\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{c}_j\|_0 \leq s, \quad c_{jj} = 0, \tag{1}$$

Introduce T auxiliary variables $\left\{\boldsymbol{b}_j^{(t)}\right\}_{t=1}^{T}$

$$\min_{\boldsymbol{c}_j, \{\boldsymbol{b}_j^{(t)}\}_{t=1}^{T}} \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{x}_j - \sum_i \xi_i^{(t)} b_{ij}^{(t)}\boldsymbol{x}_i\|_2^2$$

$$\text{s.t.} \quad \boldsymbol{b}_j^{(1)} = \cdots = \boldsymbol{b}_j^{(T)} = \boldsymbol{c}_j \quad \|\boldsymbol{b}_j^{(t)}\|_0 \leq s, \quad b_{jj}^{(t)} = 0, \quad t = 1, \cdots, T, \tag{2}$$

- To solve the consensus problem, we relax it as:

$\lambda > 0$ is a penalty parameter

$$\min_{\boldsymbol{c}_j, \{\boldsymbol{b}_j^{(t)}\}} \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{x}_j - \sum_i \xi_i^{(t)} b_{ij}^{(t)}\boldsymbol{x}_i\|_2^2 + \lambda\|\boldsymbol{b}_j^{(t)} - \boldsymbol{c}_j\|_2^2$$

$$\text{s.t.} \quad \|\boldsymbol{b}_j^{(t)}\|_0 \leq s, \quad b_{jj}^{(t)} = 0, \quad t = 1, \cdots, T, \tag{3}$$

Consensus Orthogonal Matching Pursuit (OMP-C)

➢ We develop a **Consensus Orthogonal Matching Pursuit** algorithm to solve problem:

$$\min_{\boldsymbol{c}_j, \{\boldsymbol{b}_j^{(t)}\}} \frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{x}_j - \sum_i \xi_i^{(t)} b_{ij}^{(t)} \boldsymbol{x}_i\|_2^2 + \lambda \|\boldsymbol{b}_j^{(t)} - \boldsymbol{c}_j\|_2^2$$

$$\text{s.t.} \quad \|\boldsymbol{b}_j^{(t)}\|_0 \le s, \quad b_{jj}^{(t)} = 0, \quad t = 1, \cdots, T, \tag{3}$$

by alternating the following two steps:

Each subproblem can be solved in parallel

- **Step 1**: **Fixed $\boldsymbol{c}_j$**, update $\left\{\boldsymbol{b}_j^{(t)}\right\}_{i=1}^{T}$ by solving T subproblems via **Damped OMP**
- **Step 2**: **Fixed** $\left\{\boldsymbol{b}_j^{(t)}\right\}_{i=1}^{T}$, update $\boldsymbol{c}_j$ by taking an average over $\left\{\boldsymbol{b}_j^{(t)}\right\}_{i=1}^{T}$

Damped Orthogonal Matching Pursuit (Damped OMP)

- Damped (阻尼) OMP solves:

$$\min_{\boldsymbol{b}_j} \|\boldsymbol{x}_j - \Xi\, \boldsymbol{b}_j\|_2^2 + \lambda \|\boldsymbol{b}_j - \boldsymbol{c}_j\|_2^2, \quad \text{s.t.} \quad \|\boldsymbol{b}_j\|_0 \le s, \quad b_{jj} = 0.$$

Incrementing $S^{(k)}$ by adding one index i^* at each iteration via

$$i^* = \arg \max_{i \in \mathcal{I} \setminus S^{(k)}} \psi(\boldsymbol{q}_j^{(k)}, \boldsymbol{c}_j), \quad (16)$$

where

$$\psi(\boldsymbol{q}_j^{(k)}, \boldsymbol{c}_j) := (\boldsymbol{x}_i^\top \boldsymbol{q}_j^{(k)})^2 + 2\lambda \boldsymbol{x}_i^\top \boldsymbol{q}_j^{(k)} c_{ij} - \lambda c_{ij}^2$$

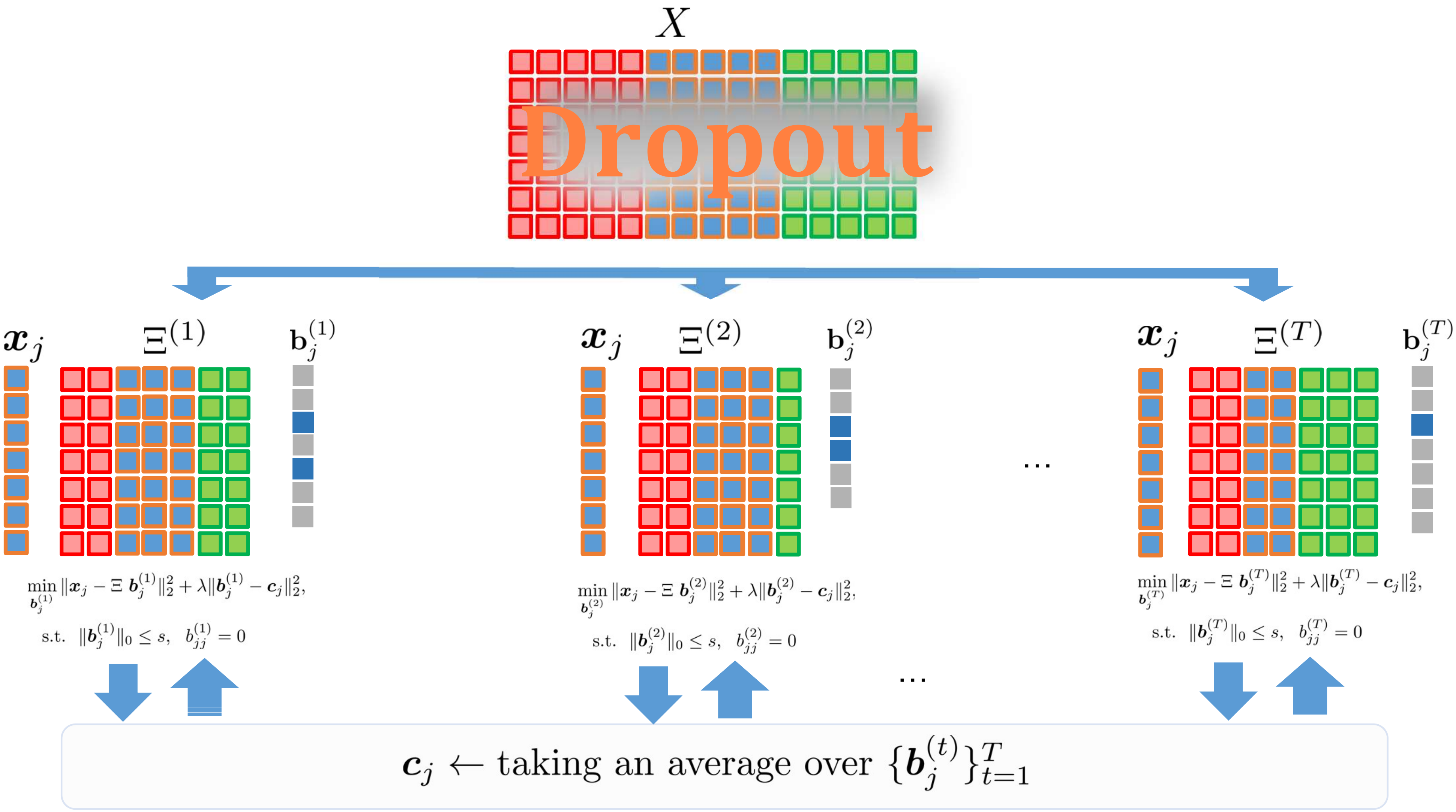
Algorithm 1 : Damped OMP

Input: Dictionary Ξ, $\mathcal{I}$, $\boldsymbol{x}_j \in I\!R^D$, $\boldsymbol{c}_j$, s, λ and ϵ.

1: Initialize $k = 0$, residual $\boldsymbol{q}_j^{(0)} = \boldsymbol{x}_j$, and $S^{(0)} = \emptyset$.
2: **while** $k < s$ and $\|\boldsymbol{q}_j^{(k)}\|_2 > \epsilon$ **do**
3: Find i^* via (16) and update $S^{(k+1)} \leftarrow S^{(k)} \bigcup \{i^*\}$;
4: Update $\boldsymbol{b}_j^{(k+1)}$ by solving (17);
5: Update $\boldsymbol{q}_j^{(k+1)} \leftarrow \boldsymbol{x}_j - \Xi\, \boldsymbol{b}_j^{(k+1)}$ and $k \leftarrow k + 1$;
6: **end while**

Output: b_j^*

$$\min_{\boldsymbol{b}_j} \|\boldsymbol{x}_j - \Xi\, \boldsymbol{b}_j\|_2^2 + \lambda \|\boldsymbol{b}_j - \boldsymbol{c}_j\|_2^2 \quad \text{s.t.} \quad \operatorname{supp}(\boldsymbol{b}_j) \subseteq S^{(k+1)}, \quad (17)$$



X
Dropout
$\boldsymbol{x}_j$
$\Xi^{(1)}$
$\mathbf{b}_j^{(1)}$
$\min_{\boldsymbol{b}_j^{(1)}} \|\boldsymbol{x}_j - \Xi\, \boldsymbol{b}_j^{(1)}\|_2^2 + \lambda\|\boldsymbol{b}_j^{(1)} - \boldsymbol{c}_j\|_2^2,$
s.t. $\|\boldsymbol{b}_j^{(1)}\|_0 \le s, \quad b_{jj}^{(1)} = 0$
$\boldsymbol{x}_j$
$\Xi^{(2)}$
$\mathbf{b}_j^{(2)}$
$\min_{\boldsymbol{b}_j^{(2)}} \|\boldsymbol{x}_j - \Xi\, \boldsymbol{b}_j^{(2)}\|_2^2 + \lambda\|\boldsymbol{b}_j^{(2)} - \boldsymbol{c}_j\|_2^2,$
s.t. $\|\boldsymbol{b}_j^{(2)}\|_0 \le s, \quad b_{jj}^{(2)} = 0$
$\cdots$
$\boldsymbol{x}_j$
$\Xi^{(T)}$
$\mathbf{b}_j^{(T)}$
$\min_{\boldsymbol{b}_j^{(T)}} \|\boldsymbol{x}_j - \Xi\, \boldsymbol{b}_j^{(T)}\|_2^2 + \lambda\|\boldsymbol{b}_j^{(T)} - \boldsymbol{c}_j\|_2^2,$
s.t. $\|\boldsymbol{b}_j^{(T)}\|_0 \le s, \quad b_{jj}^{(T)} = 0$
$\cdots$
$\boldsymbol{c}_j \leftarrow$ taking an average over $\{\boldsymbol{b}_j^{(t)}\}_{t=1}^T$

Outline

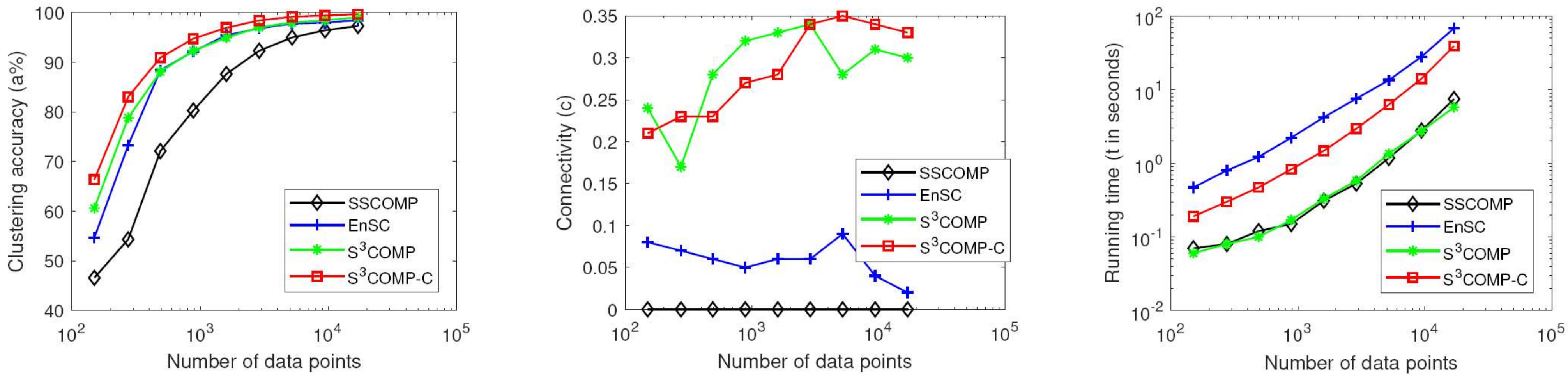
- Introduction to Subspace Clustering
- Related Work and Motivation
- Our Proposal
- Experiments
- Summary

When SSC meets Dropout …

- S^3**COMP-C**: Stochastic **Sparse Subspace Clustering via Orthogonal Matching Pursuit** with Consensus
- S^3**COMP**: Stochastic **Sparse Subspace Clustering via Orthogonal Matching Pursuit** with Consensus $(t = 1)$

Experiments on Synthetic Data: varying data points

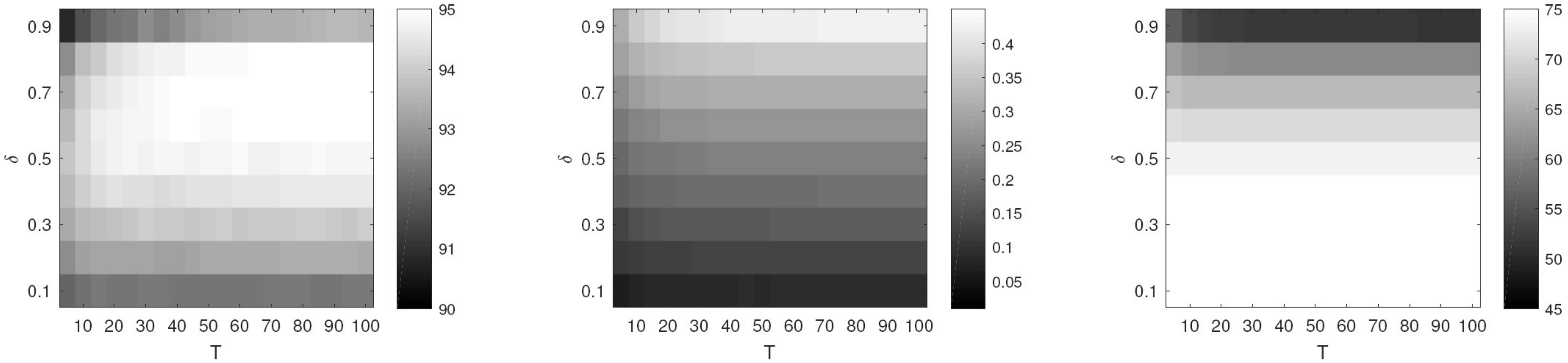
- **Synthetic Data**: Generate $n = 5$ of dimension $d = 6$ in the ambient space R^9



- Connectivity: $c = \min\left\{\lambda_2^{(i)}\right\}_{i=1}^{n}$ where $\lambda_2^{(i)}$ is the algebraic connectivity, which is defined by the second smallest eigenvalue of the normalized graph Laplacian corresponding to the i-th cluster

Experiments on Synthetic Data: varying T and δ

- **Synthetic Data**: Generate $n = 5$ of dimension $d = 6$ in the ambient space R^9

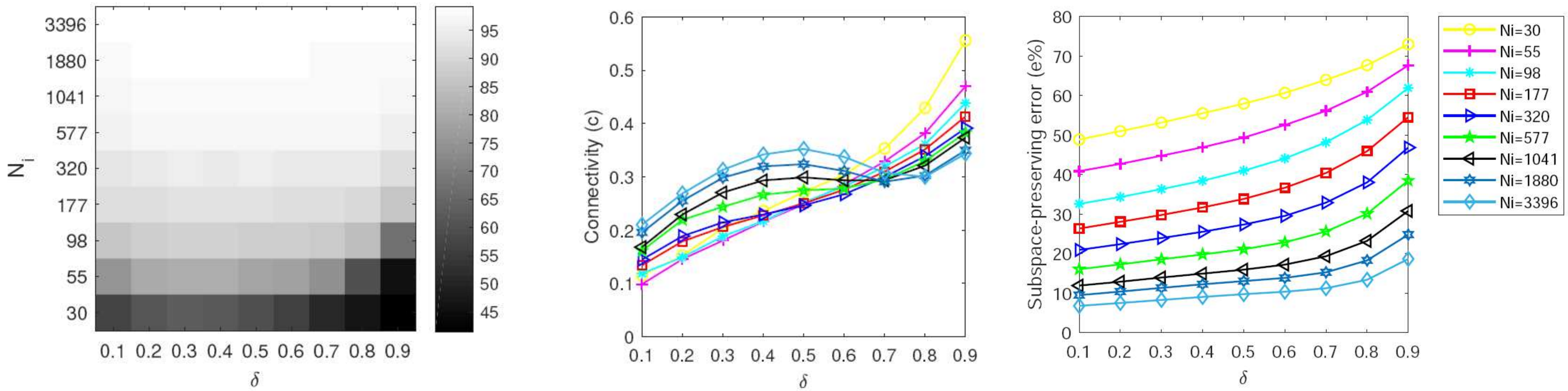


(a) Clustering accuracy ($a\%$) (b) Connectivity (c) (c) Subspace-preserving rate ($1 - e\%$)

- Experiments are conducted with N_i=320

Experiments on Synthetic Data: varying δ and N_i

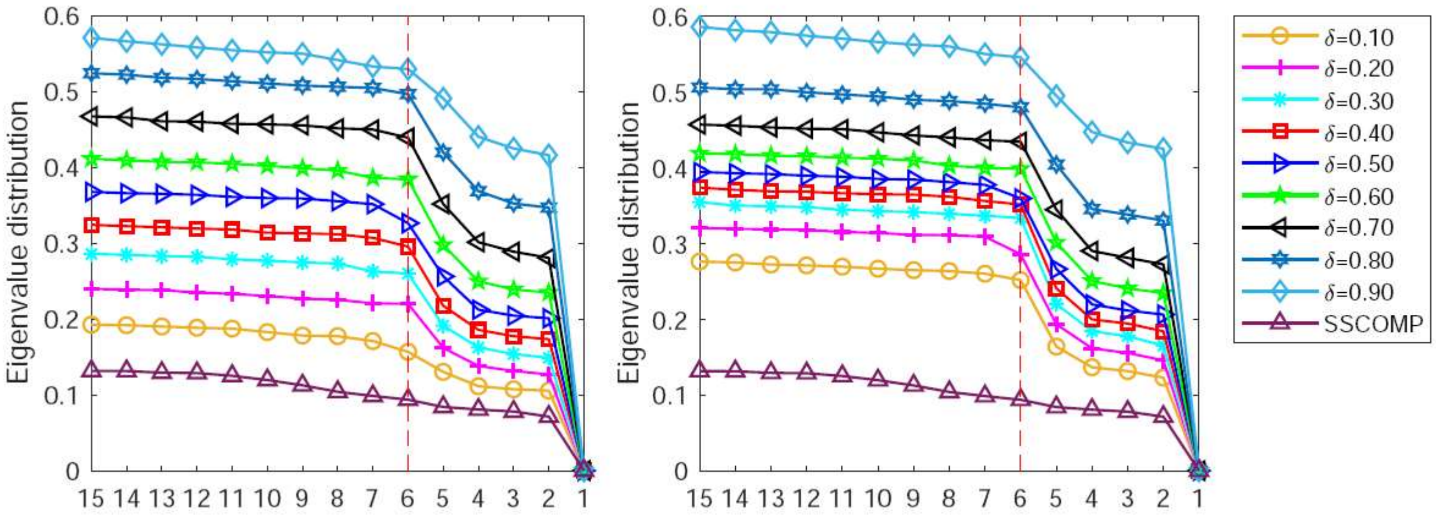
- **Synthetic Data**: Generate $n = 5$ of dimension $d = 6$ in the ambient space R^9



(a) (b) (c)

Experiments on Synthetic Data: eigenvalue gap

- **Synthetic Data**: Generate $n = 5$ of dimension $d = 6$ in the ambient space R^9

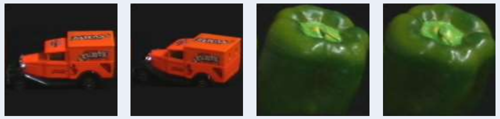


(a) S^3COMP (b) S^3COMP-C

- Experiments are conducted with N_i=320, showing the last 15 ending eigenvalues

Experiments on Real World Data

➢ Dataset Descriptions

Dataset	# data	ambient dim.	# clusters	feature	Examples
Extended Yale B	2,432	2,016	38	Raw pixel	
COIL100	7,200	1,024	100	Raw pixel	
GTSRB	12,390	500	14	HOG	
MNIST	70,000	500	10	Scattering	

Experiments on Real World Data

- Clustering accuracy compared with scalable subspace clustering methods

Dataset	# data	ESC	SR-SSC	SSCOMP	EnSC	S^3COMP-C
Extended Yale B	2,432	**87.58%**	62.11%	77.59%	61.20%	87.41%
COIL100	7,200	56.90%	58.85%	49.88%	63.94%	**78.89%**
GTSRB	12,390	90.16%	78.42%	82.52%	86.05%	**95.54%**
MNIST	70,000	90.87%	87.22%	81.59%	93.67%	**96.32%**

Experiments on Real World Data

Method	Extended Yale B			
	acc (a%)	sre (e%)	conn ($\bar{c}$)	t (sec.)
SCC	12.80	-	-	615.69
OLRSC	26.84	95.98	**0.6284**	98.25
LSR	63.99	87.57	<u>0.5067</u>	3.21
LRSC	63.17	88.75	0.4526	7.20
EnSC	61.20	23.14	0.0550	52.98
SR-SSC	62.11	-	-	79.46
SSCOMP	77.59	**20.13**	0.0381	<u>2.54</u>
ESC*	**87.58**	-	-	28.01
S^3COMP	81.61	<u>20.18</u>	0.0723	**1.92**
S^3COMP-C	<u>87.41</u>	20.28	0.0667	5.05

Experiments on Real World Data

Method	COIL100			
	acc (a%)	sre (e%)	conn ($\bar{c}$)	t(sec.)
SCC	55.24	-	-	479.13
LRSC	50.10	96.43	**0.7072**	25.11
LSR	48.22	94.95	<u>0.5246</u>	62.91
SSCOMP	49.88	14.03	0.0060	<u>13.33</u>
ESC	56.90	-	-	56.31
SR-SSC	58.85	-	-	204.38
EnSC	63.94	4.36	0.0163	19.03
S^3COMP	<u>71.47</u>	<u>3.35</u>	0.0081	**7.68**
S^3COMP-C	**78.89**	**3.15**	0.0077	20.10

Experiments on Real World Data

Method	MNIST4000 acc (a%)	MNIST4000 sre (e%)	MNIST4000 conn ($\bar{c}$)	MNIST4000 t (sec.)	MNIST10000 acc (a%)	MNIST10000 sre (e%)	MNIST10000 conn ($\bar{c}$)	MNIST10000 t (sec.)
LSR	80.02	78.53	0.6075	14.79	81.75	80.22	0.6389	147.98
LRSC	85.61	79.87	<u>0.6419</u>	4.77	89.60	81.36	<u>0.6646</u>	12.87
SCC	71.30	-	-	70.75	72.20	-	-	218.16
OLRSC	65.32	85.70	**0.8660**	47.4	67.62	86.11	**0.8738**	217.43
ESC	87.22	-	-	27.98	90.76	-	-	59.41
EnSC	85.85	**20.40**	0.1117	35.89	85.94	**16.63**	0.0938	89.21
SSCOMP	91.14	34.26	0.1371	**3.63**	93.80	32.08	0.1212	<u>11.99</u>
SR-SSC	91.70	-	-	39.24	90.05	-	-	79.87
S^3COMP	**94.30**	<u>33.15</u>	0.1529	<u>4.70</u>	<u>95.73</u>	<u>30.11</u>	0.1720	**9.14**
S^3COMP-C	<u>94.27</u>	33.26	0.1527	12.88	**95.74**	33.15	0.1719	26.50

Experiments on Real World Data

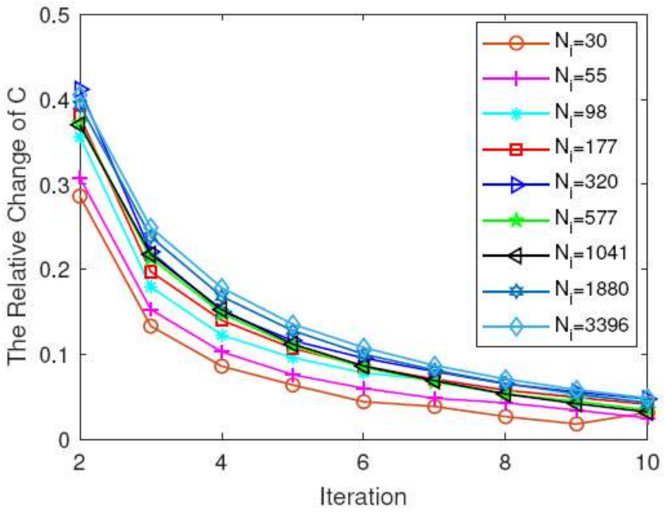
Method	MNIST70000			
	acc (a%)	sre (e%)	conn ($\bar{c}$)	t (sec.)
OLRSC	M	-	-	-
SR-SSC	87.22	-	-	585.31
SSCOMP†	81.59	<u>28.57</u>	0.0830	<u>280.58</u>
ESC	90.87	-	-	596.56
EnSC	93.67	**15.30**	0.0911	932.89
S^{3}COMP†	<u>96.31</u>	30.12	**0.1569**	**218.72**
S^{3}COMP-C†	**96.32**	30.11	**0.1569**	416.84

Experiments on Real World Data

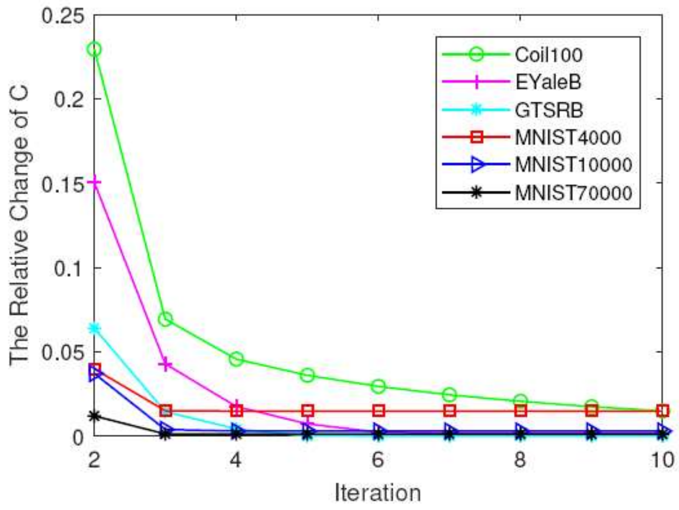
Method	GTSRB			
	acc (a%)	sre (e%)	conn ($\bar{c}$)	t (sec.)
LSR	73.93	82.80	0.6185	290.97
LRSC	87.28	78.97	<u>0.6367</u>	15.85
SCC	70.82	-	-	237.01
OLRSC	82.42	77.15	**0.7606**	291.38
SR-SSC	78.42	-	-	223.34
SSCOMP	82.52	5.42	0.0213	15.43
EnSC	86.05	**0.81**	0.0095	33.46
ESC	90.16	-	-	32.13
S^3COMP	<u>95.25</u>	<u>2.40</u>	0.0576	**3.13**
S^3COMP-C	**95.54**	2.41	0.0573	<u>7.10</u>

More Evaluations

➢ **Relative Changes of C's**



(a) Synthetic Data



(b) Real World Data

Outline

- Introduction to Subspace Clustering
- Related Work and Motivation
- Our Proposal
- Experiments
- Summary

Summary

- **Self-Expression Model + Dropout**
  - Equivalent to adding an **implicit squared ℓ_2 norm** regularization

- **Stochastic Sparse Subspace Clustering**

When SSC meets Dropout:

- **S^3COMP-C**: **Stochastic Sparse Subspace Clustering via Orthogonal Matching Pursuit** with **Consensus**
  - ✓ Improved connectivity
  - ✓ Better clustering accuracy
  - ✓ More flexible scalability: parallel & efficient

Acknowledgement

- This work is supported by:
  - NSFC under grant 61876022
  - Open Project Fund from Key Laboratory of Machine Perception (MOE), Peking University

Thank you for your attention!