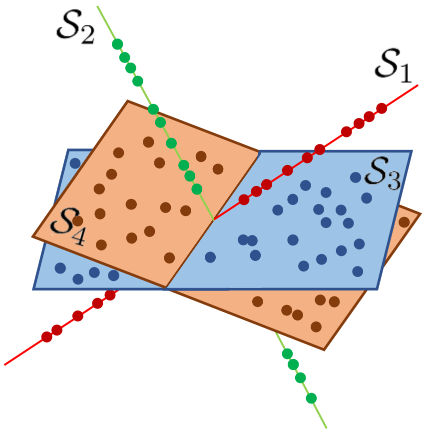
Scalable Direction-Search-Based Approach to Subspace Clustering

Yicong He and George K. Atia
Department of Electrical and Computer Engineering, University of Central Florida

Yicong.He@ucf.edu, George.Atia@ucf.edu

UCF

Subspace clustering



Data **X** lies in a union of low-dimensional subspaces $\cup_{i=1}^{k}\mathcal{S}_i$

The goal of subspace clustering : Grouping data belonging to the same subspace (cluster).

Representation-based subspace clustering

$$\min_{\mathbf{C}} \|\mathbf{X}-\mathbf{XC}\|_F^2 + f(\mathbf{C}), \ \text{s.t. } \operatorname{diag}(\mathbf{C}) = 0$$

$f(\mathbf{C}) = \|\mathbf{C}\|_1$ Sparse Subspace Clustering (SSC) [Elhamifar and Vidal, 2009]

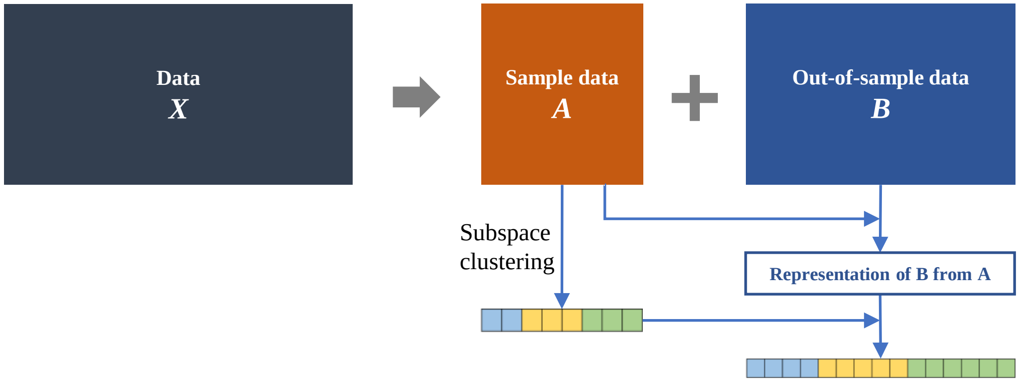
$f(\mathbf{C}) = \|\mathbf{C}\|_*$ Low Rank Representation (LRR) [Liu et al., 2012]

$f(\mathbf{C}) = \|\mathbf{C}\|_F^2$ Least Square Regression (LSR) [Lu et al., 2012]

Find self-representation on the whole data ***whole data***

Large-scale Subspace clustering

Scalable Framework for Representation-based subspace clustering [Peng et al., 2016]



Elastic Net Subspace Clustering (EnSC) [You et al., 2015]

Landmark-based Clustering (LSC) [Cai and Chen, 2015]

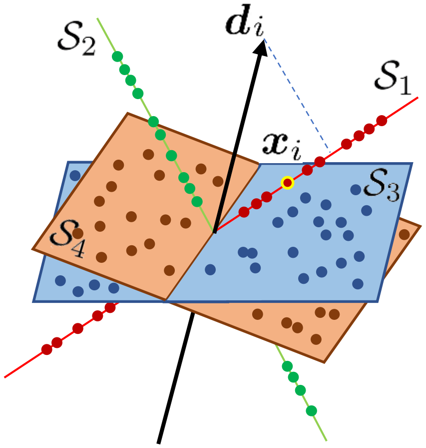
Scalable and robust SSC (SR-SSC) [Abdolali et al., 2019]

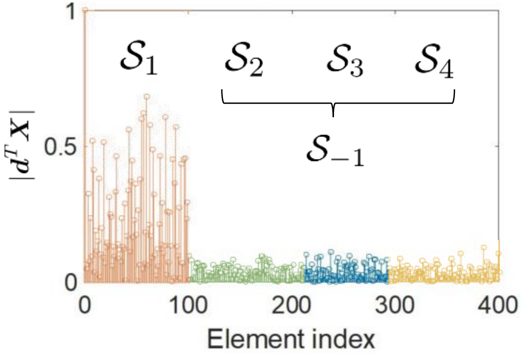
Find self-representation on ***a small subset of data***

Optimal Direction Search [Rahmani and Atia, 2017]

$$\min_{\boldsymbol{d}} \|\boldsymbol{d}^T \boldsymbol{X}\|_p \quad \text{s.t.} \quad \boldsymbol{d}^T \boldsymbol{x}_i = 1$$

It searches for an optimal direction $\boldsymbol{d}_i \in \mathbb{R}^{M \times 1}$ for each data point $\boldsymbol{x}_i$ by minimizing its projections on all data points except $\boldsymbol{x}_i$





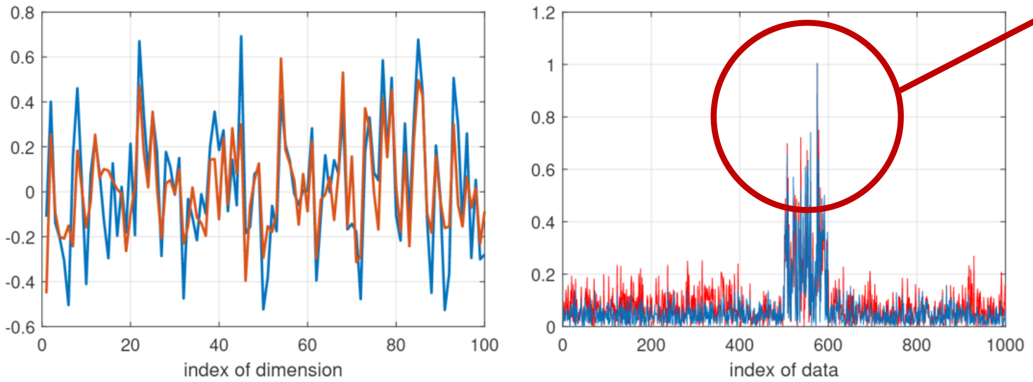
Large coherence with data in the same subspace

Find directions on the ***whole data***

Do we need to compute the direction using all the data?

Use subset $\hat{\boldsymbol{X}}$ $\quad \hat{\boldsymbol{d}}_i = \arg\min_{\boldsymbol{d}} \|\boldsymbol{d}^T \hat{\boldsymbol{X}}\|_p \quad \text{s.t.} \quad \boldsymbol{d}^T \hat{\boldsymbol{x}}_i = 1$

Coherence is highly preserved



Computed $\boldsymbol{d}_i, \widehat{\boldsymbol{d}}_i$ (left) and corresponding $|\boldsymbol{d}_i^T X|$, $|\widehat{\boldsymbol{d}}_i^T X|$ (right) using full data (blue) and 20% of the data (red)

Find directions on ***a small subset of data***

Also, the direction search is friendly to out-of-sample data…

For a data point $\boldsymbol{b}_i$, its cluster label can be obtained using $|\boldsymbol{b}_i^T \boldsymbol{D}|$

Theoretical analysis

Proposition.1 *Defining* $\boldsymbol{d}_{OP} = \arg \min_{\substack{\boldsymbol{d}^T\boldsymbol{x}=1 \\ \boldsymbol{x} \in \mathcal{S}_j}} \|\boldsymbol{d}^T\boldsymbol{X}\|_1$, *if the following conditions are satisfied*

(1) $\sum_{\boldsymbol{x}_i \in \boldsymbol{X}^{-j}} |\boldsymbol{v}^T\boldsymbol{x}_i| + \sum_{\substack{\boldsymbol{x}_i \in \boldsymbol{X}^j \\ i \in I_0}} |\boldsymbol{v}^T\boldsymbol{x}_i| > \sum_{\boldsymbol{x}_i \in \boldsymbol{X}^j} \mathrm{sgn}(\boldsymbol{x}_i^T\boldsymbol{d}_{OP})\boldsymbol{v}^T\boldsymbol{x}_i$

$I_0 = \{i \in [n_j] : \boldsymbol{x}_i^T\boldsymbol{d}_{op} = 0, \boldsymbol{x}_i \in \boldsymbol{X}^j\}$

$\boldsymbol{v}$—an arbitrarily small vector satisfying $\boldsymbol{v}^T\boldsymbol{x} = 0$

(2) $\mathcal{S}_j \nsubseteq \mathcal{S}_{-j}, j \in [k]$

(3) $M \gg r_j, j \in [k]$ $\quad$ M—ambient dimension $\quad$ r_j—dimension of $\mathcal{S}_j$

Then we have $\boldsymbol{d}_{OP} \in \mathcal{S}_{-j}^{\perp} \Longleftrightarrow \boxed{\boldsymbol{d}_{OP}^T\boldsymbol{X}^{-j} = \boldsymbol{0}}$

According to Theorem 1, for a subset $\hat{\boldsymbol{X}}$, we have

$$\hat{\boldsymbol{d}}_{OP} = \arg \min_{\substack{\boldsymbol{d}^T\boldsymbol{x}=1 \\ \boldsymbol{x} \in \mathcal{S}_j}} \|\boldsymbol{d}^T\hat{\boldsymbol{X}}\|_1 \Longrightarrow \hat{\boldsymbol{d}}_{OP}^T\hat{\boldsymbol{X}}^{-j} = \boldsymbol{0} \Longrightarrow \boxed{\hat{\boldsymbol{d}}_{OP}^T\boldsymbol{X}^{-j} = \hat{\boldsymbol{d}}_{OP}^T\hat{\boldsymbol{X}}^{-j}\boldsymbol{C}^{-j} = \boldsymbol{0}}$$

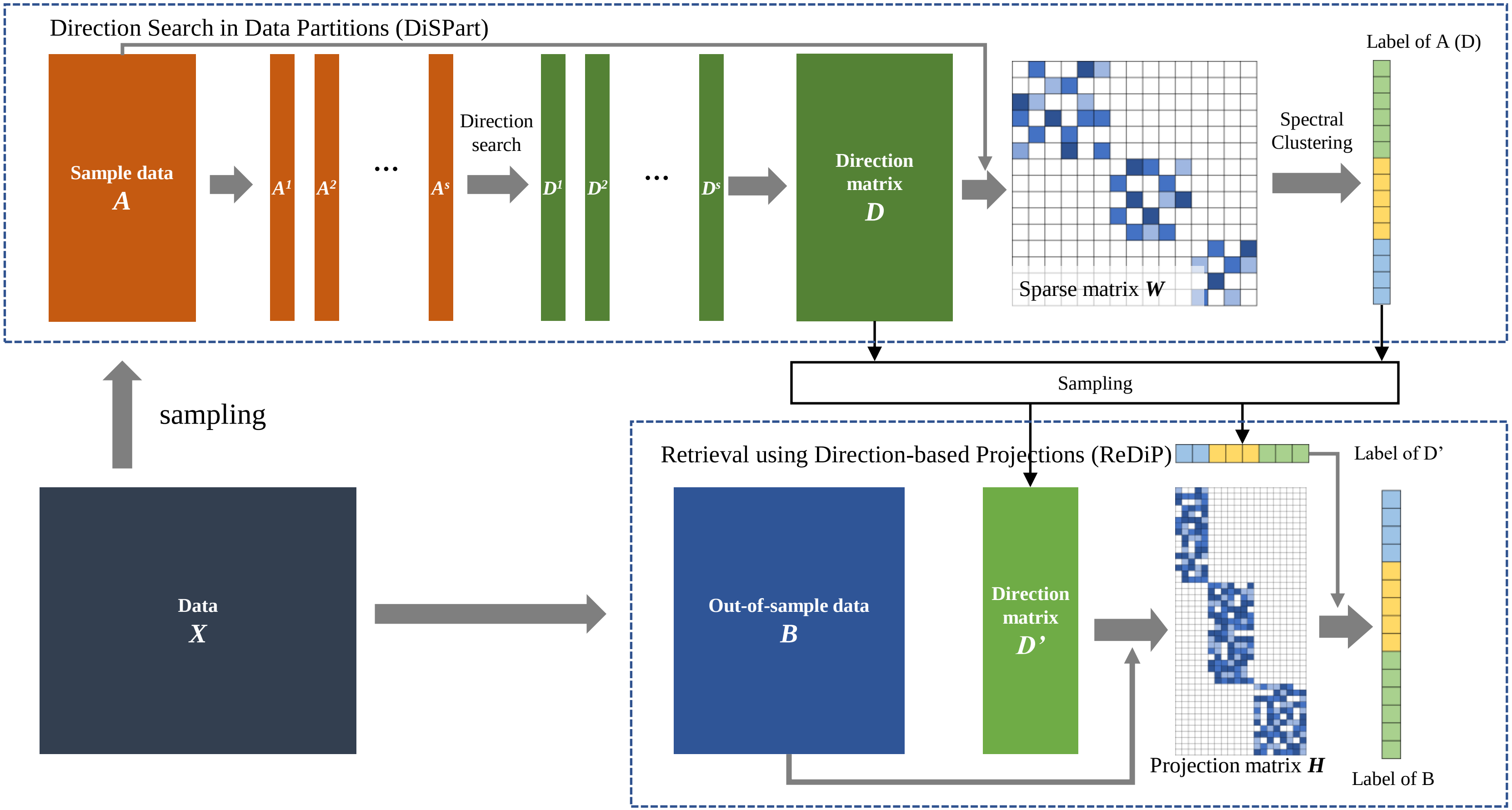
$\boldsymbol{X}^{-j} = \hat{\boldsymbol{X}}^{-j}\boldsymbol{C}^{-j}$

self-representation

We can find correct neighborhoods with $|\hat{\boldsymbol{d}}_{OP}^T\boldsymbol{X}|$

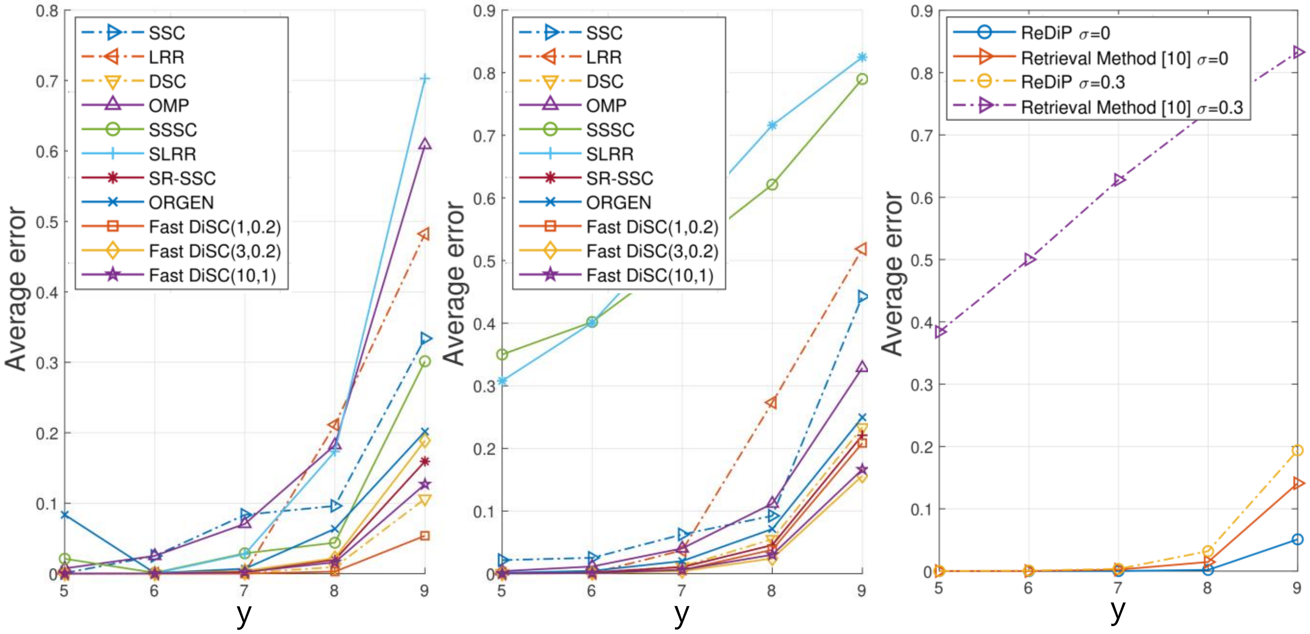
UCF

Framework of Fast Direction Search-Based Subspace Clustering (Fast DiSC)



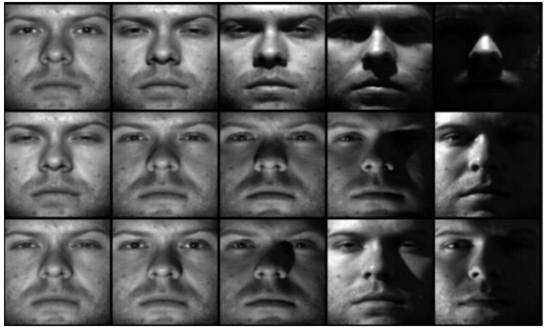
Experiment on synthetic data

$\mathbf{X}$ lies in $\cup_{i=1}^{k} \mathcal{S}_i \quad \mathcal{S}_i = \mathcal{R}_i \oplus \boxed{\mathcal{M}}$ *y*-dimensional subspace shared by all subspaces



Average clustering error as function of the dimension of the intersection y. Left: noiseless (σ = 0) on whole process. Middle: noisy (σ = 0.3) on whole process. Right: Retrieval performance comparison only.

Experiment on real datasets



Extended Yale B

38 subjects
2432 images

USPS

10 subjects
9298 images

MNIST

10 subjects
70000 images

Experiment results on real Datasets

Dataset	Evaluation	OMP	SSSC	SLRR	SR-SSC	ORGEN	Fast DiSC (s_1,0.2)	Fast DiSC (s_2,0.5)	Fast DiSC (s_3,1)	SSC	LRR	DSC
ExYale B	Acc(%)	71.90	59.67	56.43	71.03	60.90	50.48	78.00	**90.30**	67.59	73.43	**91.89**
	NMI(%)	79.31	66.22	67.60	76.84	70.95	58.97	79.82	**92.20**	73.08	83.11	**93.53**
	Time(s)	11.59	25.96	32.87	35.68	44.39	4.36	19.87	29.86	124.67	66.06	77.34
USPS	Acc(%)	19.77	45.18	47.82	66.62	48.92	70.01	74.17	**77.53**	48.56	55.77	**64.70**
	NMI(%)	6.87	52.47	52.48	68.02	56.70	69.77	73.11	**78.24**	58.77	61.79	**78.59**
	Time(s)	11.52	71.41	22.96	76.19	78.70	4.11	6.88	13.31	3881.29	385.80	1379.79
MNIST	Acc(%)	46.26	85.08	83.75	83.63	93.79	89.82	87.89	**98.03**	-	-	-
	NMI(%)	51.55	87.33	87.13	83.38	88.80	86.37	87.42	**94.67**	-	-	-
	Time(s)	695.41	713.47	194.37	1265.06	959.22	107.36	270.29	705.82	-	-	-

Fast DiSC(s,θ): s-number of partitions θ-sampling rate (i.e., $N_1 = \lfloor \theta N \rfloor$) in DiSPart

Summary of contribution of the work

(1)**We develop Fast DiSC for large-scale subspace clustering**, which consists of two newly proposed procedures DiSPart and ReDiP.

(2)**We provide theoretical analysis for direction search**, showing that a small part of the data generally suffices to identify correct neighborhood sets for the data points in their corresponding subspaces.

(3)**We present experimental results with comparisons to state-of-the-art algorithms on both synthetic and real data**, demonstrating the superior performance of the Fast DiSC.

Thank you!