

# Bi-Sparse Unsupervised Feature Selection

Xianchao Xiu

Department of Automation



Operations Research Society of China, Oct 18-21, 2024

Joint work with [Chenyi Huang](#) (SHU), [Pan Shang](#) (CAS) and [Wanquan Liu](#) (SYSU)

# Outline

Introduction

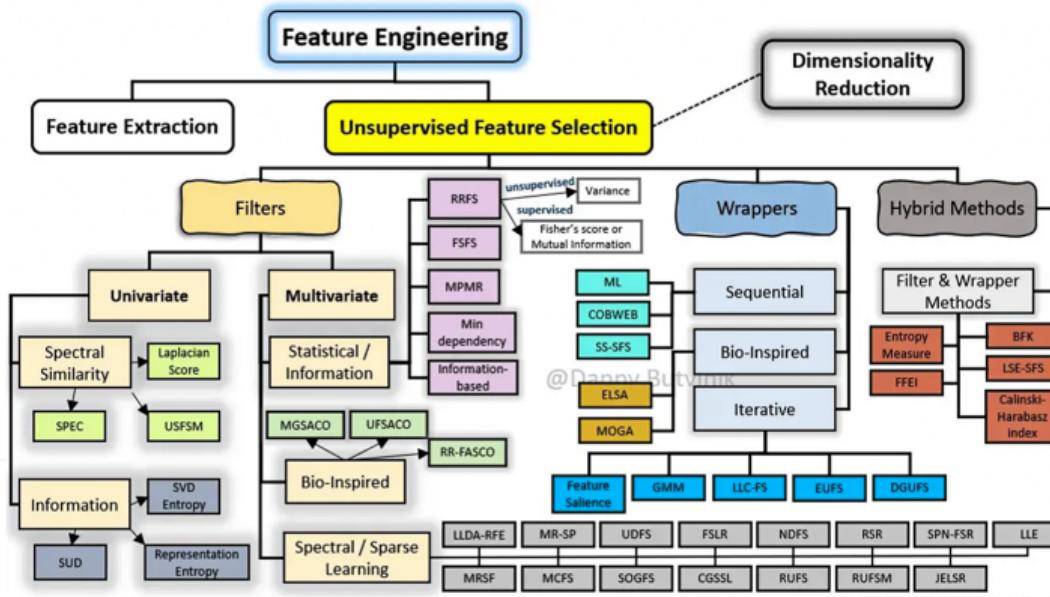
Proposed Method

Numerical Experiments

Conclusions and Future Work

# Unsupervised Feature Selection

- ▶ Select a subset of input features without labels
- ▶ <https://dannybutvinik.medium.com>



## Related Works

- ▶ Yang-Shen-Huang-Zhou, IJCAI, 2011

$$\begin{aligned} \min_W & -\text{Tr}(W^\top SW) + \lambda \|W\|_{2,1} \\ \text{s.t. } & W^\top W = I \end{aligned}$$

- ▶ Tian-Nie-Wang-Li, NIPS, 2020

$$\begin{aligned} \min_W & -\text{Tr}(W^\top SW) + \lambda \|W\|_{2,0} \\ \text{s.t. } & W^\top W = I \end{aligned}$$

- ▶ Li-Nie-Bian-Wu-Li, IEEE TPAMI, 2023

$$\begin{aligned} \min_W & -\text{Tr}(W^\top SW) + \lambda \|W\|_{2,p}^p \quad (0 < p < 1) \\ \text{s.t. } & W^\top W = I \end{aligned}$$

## Related Works

- ▶ Zhu-Zhang-Wen-He-Cheng, MTA, 2017

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^\top SW) + \lambda_1 \|W\|_{2,1} + \lambda_2 \|W\|_1 \\ \text{s.t. } & W^\top W = I \end{aligned}$$

- ▶ Other fields

- ▶ Rubinstein-Zibulevsky-Huang-Elad, IEEE TSP, 2010
- ▶ Hu-Liu-Gao-Shang, IEEE TCBB, 2021
- ▶ Bian-Xu-Wang, IEEE PIMRC, 2022
- ▶ Zhang-Liu-Li, IEEE TIP, 2023

**Can non-convex bi-sparse optimization benefit UFS?**

# Outline

Introduction

Proposed Method

Numerical Experiments

Conclusions and Future Work

## New Formulation

- Bi-sparse unsupervised feature selection

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^\top SW) + \lambda \|W\|_{2,p}^p \quad (0 < p < 1) \\ \text{s.t. } \quad & W^\top W = I \end{aligned}$$

↓

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^\top SW) + \lambda_1 \|W\|_{2,p}^p + \lambda_2 \|W\|_q^q \quad (0 \leq p, q < 1) \\ \text{s.t. } \quad & W^\top W = I \end{aligned}$$

- Advantages

- Bi-sparse optimization:  $\lambda_1 \|W\|_{2,p}^p + \lambda_2 \|W\|_q^q$
- A non-convex framework:  $0 \leq p, q < 1$

# Optimization Algorithm

- ▶ First-order algorithm: PAM (Proximal Alternating Method)
- ▶ Model reformulation

$$\min_W -\text{Tr}(W^\top SW) + \lambda_1 \|W\|_{2,p}^p + \lambda_2 \|W\|_q^q$$

$$\text{s.t. } W^\top W = I$$

$\Downarrow$

$$\min_{W,U,V} -\text{Tr}(W^\top SW) + \lambda_1 \|V\|_{2,p}^p + \lambda_2 \|U\|_q^q$$

$$\text{s.t. } W^\top W = I, V = W, U = W$$

$\Downarrow$

$$\begin{aligned} \min_{W,U,V} & -\text{Tr}(W^\top SW) + \lambda_1 \|V\|_{2,p}^p + \lambda_2 \|U\|_q^q \\ & + \frac{\beta_1}{2} \|W - U\|_F^2 + \frac{\beta_2}{2} \|W - V\|_F^2 + \Phi(W) \end{aligned}$$

# Optimization Algorithm

- ▶ Input:  $X, \lambda_1, \lambda_2, \beta_1, \beta_2, p, q, \tau_1, \tau_2, \tau_3$
- ▶ Initialize:  $W^0, U^0, V^0$
- ▶ While not converged do
  - ▶ Update  $W^{k+1}$  by

$$\begin{aligned} \min_W \quad & -\text{Tr}(W^\top SW) + \frac{\beta_1}{2} \|W - U^k\|_F^2 + \frac{\beta_2}{2} \|W - V^k\|_F^2 + \frac{\tau_1}{2} \|W - W^k\|_F^2 \\ \text{s.t. } & W^\top W = I \end{aligned}$$

- ▶ Update  $U^{k+1}$  by

$$\min_U \lambda_2 \|U\|_q^q + \frac{\beta_1}{2} \|W^{k+1} - U\|_F^2 + \frac{\tau_2}{2} \|U - U^k\|_F^2$$

- ▶ Update  $V^{k+1}$  by

$$\min_V \lambda_1 \|V\|_{2,p}^p + \frac{\beta_2}{2} \|W^{k+1} - V\|_F^2 + \frac{\tau_3}{2} \|V - V^k\|_F^2$$

- ▶ Output:  $W^{k+1}, U^{k+1}, V^{k+1}$

# Update $W$

- ▶ Riemannian gradient

$$\min_W -\text{Tr}(W^\top SW) + \frac{\beta_1}{2} \|W - U^k\|_{\text{F}}^2 + \frac{\beta_2}{2} \|W - V^k\|_{\text{F}}^2 + \frac{\tau_1}{2} \|W - W^k\|_{\text{F}}^2$$

$$\text{s.t. } W^\top W = I$$



$$\nabla g(W) = -2SW + \beta_1(W - U^k) + \beta_2(W - V^k) + \tau_1(W - W^k)$$



$$\begin{aligned}\text{grad } g(W) &= \mathcal{P}_W(\nabla g(W)) \\ &= \nabla g(W) - W \text{sym}(W^\top \nabla g(W))\end{aligned}$$

# Update $W$

## ► Riemannian Hessian

$$\nabla^2 g(W) = -2I \otimes S + (\beta_1 + \beta_2 + \tau_1)I$$

$\Downarrow$

$$\begin{aligned}\text{Hess } g(W) &= \mathcal{P}_W(\nabla^2 g(W)) \\ &= \nabla^2 g(W) - W \text{sym}(W^\top \nabla^2 g(W))\end{aligned}$$

$\Downarrow$

$$\text{Hess } g(W) \approx \frac{\text{grad } g(W + \varepsilon I) - \text{grad } g(W)}{\varepsilon}$$

## Update $W$

- ▶ Input:  $S, U^k, V^k, \beta_1, \beta_2, \tau_1, \varepsilon, \Delta' > 0, \rho' \in [0, \frac{1}{4})$
- ▶ While not converged do
  - ▶ Obtain  $\eta_i$  by solving trust domain subproblem

$$\begin{aligned} \min_{\eta \in T_W \text{St}(d,m)} m_W(\eta) &= g(W) + \text{Tr}(\eta^\top \text{grad } g(W)) + \frac{1}{2} \text{vec}(\eta)^\top \text{Hess } g(W) \text{vec}(\eta) \\ \text{s.t.} \quad \text{Tr}(\eta^\top W \eta^\top) &\leq \Delta^2 \end{aligned}$$

- ▶ Compute the trust ratio  $\rho_i$
  - ▶ if  $\rho_i < \frac{1}{4}$  then
$$\Delta_{i+1} = \frac{1}{4}\Delta_i$$
  - ▶ else if  $\rho_i > \frac{3}{4}$  and  $\|\eta_i\| = \Delta_i$  then
$$\Delta_{i+1} = \min(2\Delta_i, \Delta')$$
  - ▶ else
$$\Delta_{i+1} = \Delta_i$$
  - ▶ if  $\rho_i > \rho'$  then
$$W_{i+1}^k = R_W(\eta_i)$$
  - ▶ else
$$W_{i+1}^k = W_i^k$$
- ▶ Output:  $W$

## Update $U$

$$\min_U \lambda_2 \|U\|_q^q + \frac{\beta_1}{2} \|W^{k+1} - U\|_{\text{F}}^2 + \frac{\tau_2}{2} \|U - U^k\|_{\text{F}}^2$$

↓

$$\min_U \lambda_2 \|U\|_q^q + \frac{\beta_1 + \tau_2}{2} \|U - \frac{\beta_1}{\beta_1 + \tau_2} W^{k+1} + \frac{\tau_2}{\beta_1 + \tau_2} U^k\|_{\text{F}}^2$$

↓

$$\min_{u_{ij}} \lambda_2 |u_{ij}|^q + \frac{\beta_1 + \tau_2}{2} (u_{ij} - y_{ij})^2$$

↓

$$u_{ij} = \text{Prox}(y_{ij}, \lambda_2 / (\beta_1 + \tau_2))$$

## Lemma

- Revisiting  $\ell_q$  ( $0 \leq q < 1$ ) Norm Regularized Optimization, arXiv:2306.14394

$$\begin{aligned}\text{Prox}(a, \lambda) &= \operatorname{argmin}_x \frac{1}{2}(x - a)^2 + \lambda|x|^q \quad (0 \leq q < 1) \\ &= \begin{cases} \{0\}, & |a| < \kappa(\lambda, q) \\ \{0, \operatorname{sgn}(a)c(\lambda, q)\}, & |a| = \kappa(\lambda, q) \\ \{\operatorname{sgn}(a)\varpi_q(|a|)\}, & |a| > \kappa(\lambda, q) \end{cases}\end{aligned}$$

where

$$c(\lambda, q) = (2\lambda(1-q))^{\frac{1}{2-q}} > 0$$

$$\kappa(\lambda, q) = (2-q)\lambda^{\frac{1}{2-q}}(2(1-q))^{\frac{q+1}{q-2}}$$

$$\varpi_q(a) \in \{x : x - a + \lambda q \operatorname{sgn}(x)x^{q-1} = 0, x > 0\}$$

## Update $V$

$$\min_V \lambda_1 \|V\|_{2,p}^p + \frac{\beta_2}{2} \|W^{k+1} - V\|_{\text{F}}^2 + \frac{\tau_3}{2} \|V - V^k\|_{\text{F}}^2$$

$\Downarrow$

$$\min_V \lambda_1 \|V\|_{2,p}^p + \frac{\beta_2 + \tau_3}{2} \|V - \frac{\beta_2}{\beta_2 + \tau_3} W^{k+1} + \frac{\tau_3}{\beta_2 + \tau_3} V^k\|_{\text{F}}^2$$

$\Downarrow$

$$\min_{\mathbf{v}^i} \lambda_1 \sum_{i=1}^d \|\mathbf{v}^i\|_2^p + \frac{\beta_2 + \tau_3}{2} \|\mathbf{v}^i - \mathbf{z}^i\|_2^2$$

$\Downarrow$

$$\mathbf{v}^i = \text{Prox}(\|\mathbf{z}^i\|_2, \lambda_1 / (\beta_2 + \tau_3)) \cdot \frac{\mathbf{z}^i}{\|\mathbf{z}^i\|_2}$$

# Outline

Introduction

Proposed Method

Numerical Experiments

Conclusions and Future Work

# Experimental Details

- ▶ Compared methods
  - ▶ LapScore: He-Cai-Niyogi, NIPS, 2005
  - ▶ MCFS: Cai-Zhang-He, SIGKDD, 2010
  - ▶ UDFS: Yang-Shen-Ma, IJCAI, 2011
  - ▶ SOGFS: Nie-Zhu-Li, IEEE TKDE, 2021
  - ▶ RNE: Liu-Ye-Li-Wang, KBS, 2020
  - ▶ FSPCA: Tian-Nie-Wang-Li, NIPS, 2020
  - ▶ SPCAFS: Li-Nie-Bian, IEEE TPAMI, 2023
  - ▶ GSPCA: Zhu-Zhang-Wen, MTA, 2017
- ▶ Implementation setups
  - ▶ Initialization: QR decomposition
  - ▶ Stopping criteria:

$$\frac{|f(W^{k+1}, U^{k+1}, V^{k+1}) - f(W^k, U^k, V^k)|}{\max\{1, |f(W^k, U^k, V^k)|\}} \leq 10^{-4}$$

# Experimental Details

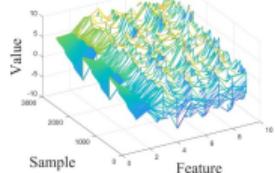
- ▶ Selected datasets

Type	Datasets	Features	Samples	Classes
Synthetic datasets	Dartboard1	9	1000	4
	Diamond9	9	3000	9
Real-world datasets	COIL20	1024	1440	20
	USPS	256	1000	10
	LUNG	325	73	7
	GLIOMA	4434	50	4
	UMIST	644	575	20
	pie	1024	1166	53
	Isolet	617	1560	26
	MSTAR	1024	2425	10

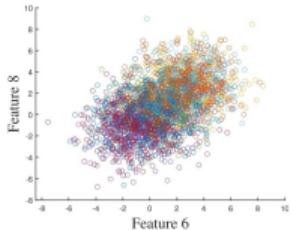
- ▶ Evaluation metrics

- ▶ ACC: Accuracy
- ▶ NMI: Normalized mutual information

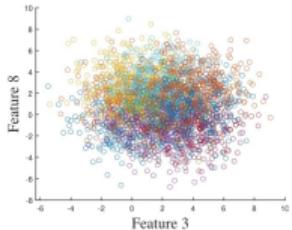
# Synthetic Experiments



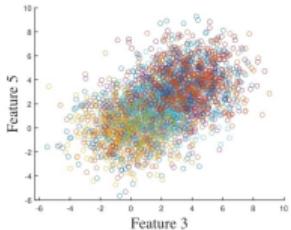
(a) Diamond9



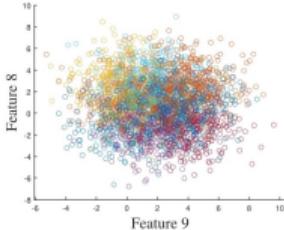
(b) LapScore



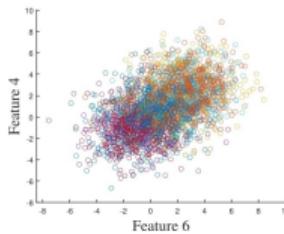
(c) MCFS



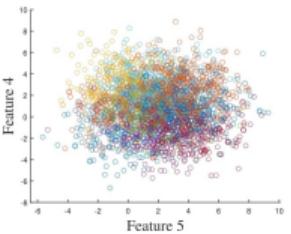
(d) SOGFS



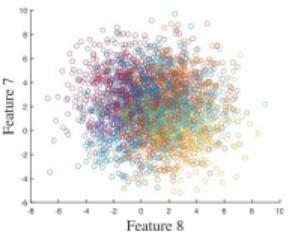
(e) RNE



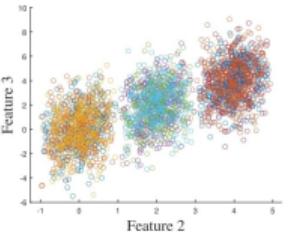
(f) UDFS



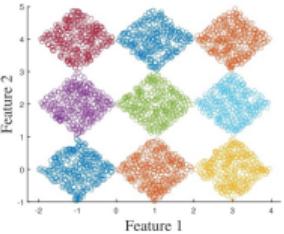
(g) SPCAFS



(h) FSPCA



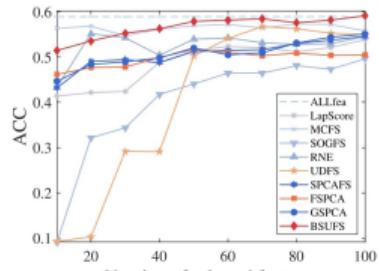
(i) GSPCA



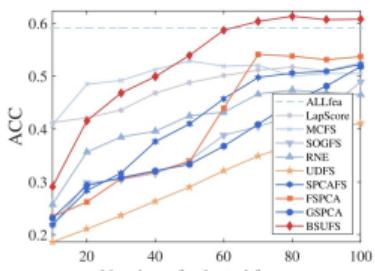
(j) BSUFS

# Real Experiments

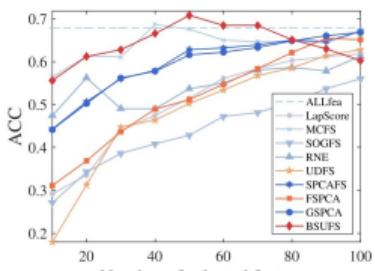
## ► ACC comparisons



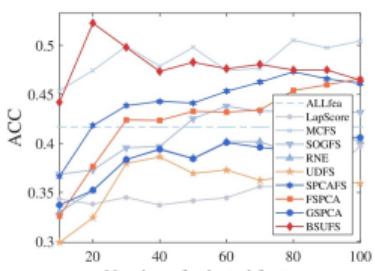
(a) COIL20



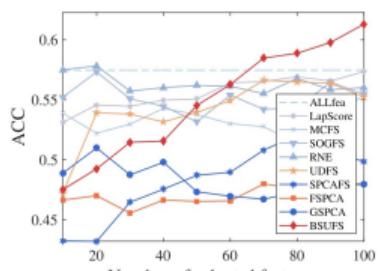
(b) Isolet



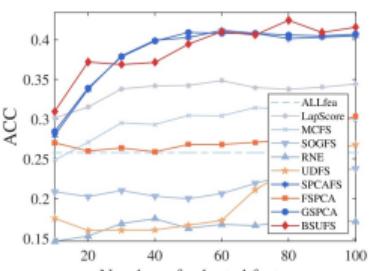
(c) USPS



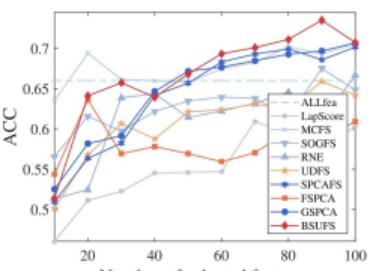
(d) umist



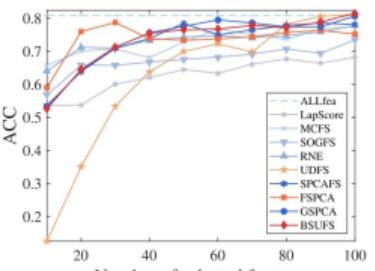
(e) GLIOMA



(f) pie



(g) LUNG



(h) MSTAR

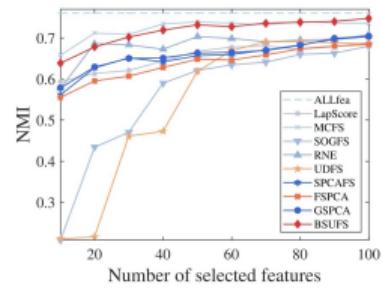
# Real Experiments

## ► ACC comparisons

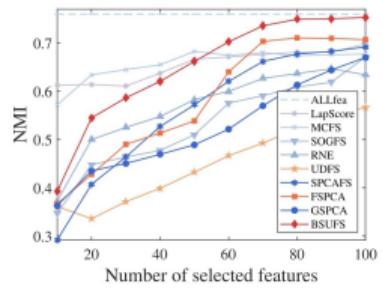
Datasets	ALLfea	LapScore	MCFS	SOGFS	RNE	UDFS	SPCAFS	FSPCA	GSPCA	BSUFS
COIL20	58.97±4.99 (10)	53.91±3.61 (100)	<b>57.35±3.83 (80)</b>	49.66±3.63 (100)	55.16±3.35 (20)	56.77±3.09 (70)	54.63±3.64 (100)	51.71±3.05 (50)	55.12±2.67 (100)	<b>59.18±3.49 (100)</b>
Isolet	59.18±3.19 (10)	52.55±2.83 (100)	52.87±2.87 (50)	48.93±2.69 (100)	47.39±2.91 (80)	41.11±1.71 (100)	52.26±2.81 (100)	<b>54.15±2.69 (70)</b>	51.84±2.82 (100)	<b>61.34±3.33 (80)</b>
USPS	67.79±4.96 (10)	61.76±4.52 (100)	<b>68.70±4.10 (40)</b>	56.00±3.48 (100)	61.28±3.46 (100)	62.83±3.79 (100)	66.98±3.92 (100)	65.43±4.90 (90)	66.79±4.10 (100)	<b>70.77±3.73 (50)</b>
umist	41.68±2.46 (10)	39.71±3.28 (100)	<b>50.54±4.16 (80)</b>	43.81±2.98 (60)	41.01±2.25 (90)	38.64±1.61 (40)	47.32±3.48 (80)	46.58±2.34 (100)	40.65±2.29 (90)	<b>52.29±3.61 (20)</b>
GLIOMA	57.44±6.40 (10)	57.36±3.60 (100)	55.52±9.25 (100)	57.32±6.47 (20)	<b>57.80±2.98 (20)</b>	56.64±6.47 (70)	52.08±3.64 (80)	48.04±5.26 (90)	51.00±5.08 (20)	<b>61.28±9.01 (100)</b>
pie	25.79±1.39 (10)	34.86±1.43 (60)	31.46±1.47 (70)	23.78±1.19 (100)	17.49±0.76 (40)	26.82±1.32 (100)	<b>41.16±2.46 (60)</b>	30.39±1.43 (100)	40.90±1.85 (50)	<b>42.45±1.74 (80)</b>
LUNG	66.03±7.23 (10)	60.93±8.02 (70)	70.55±7.66 (100)	67.53±7.73 (90)	66.68±8.32 (100)	65.89±7.43 (90)	70.16±7.71 (100)	63.62±5.45 (20)	<b>70.68±7.41 (100)</b>	<b>73.51±6.80 (90)</b>
MSTAR	80.81±8.76 (10)	68.21±4.57 (100)	77.60±8.32 (100)	73.46±5.61 (100)	77.82±6.16 (100)	<b>81.25±7.48 (100)</b>	78.63±8.68 (90)	78.74±5.20 (30)	80.65±6.47 (100)	<b>81.43±6.89 (100)</b>
Average	57.21±4.92	53.66±3.98	<b>58.07±5.21</b>	52.56±4.22	53.08±3.77	53.74±4.11	57.90±4.54	54.83±3.79	57.21±4.09	<b>62.78±4.83</b>

# Real Experiments

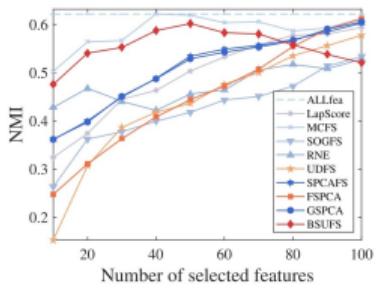
## ► NMI comparisons



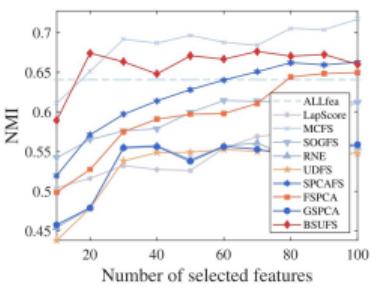
(a) COIL20



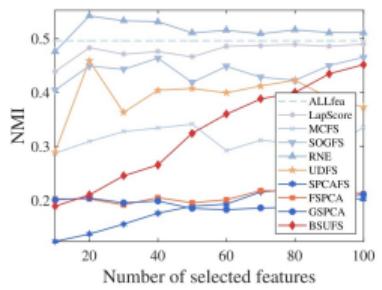
(b) Isolet



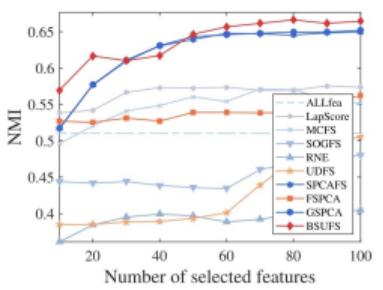
(c) USPS



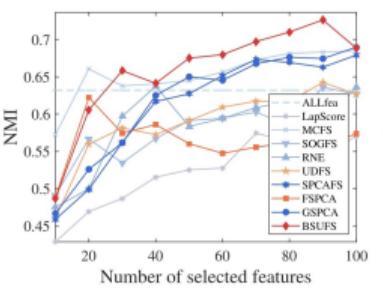
(d) umist



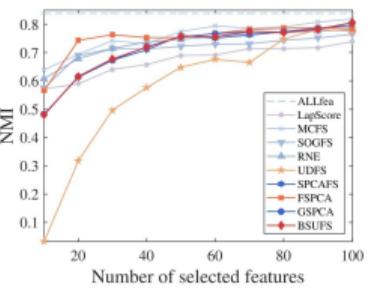
(e) GLIOMA



(f) pie



(g) LUNG



(h) MSTAR

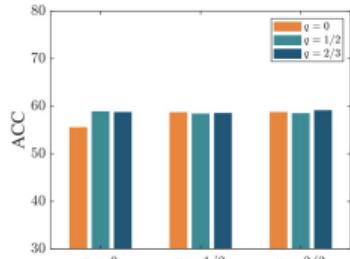
# Real Experiments

## ► NMI comparisons

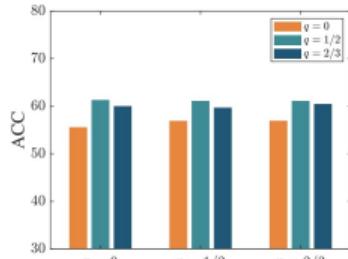
Datasets	ALLfea	LapScore	MCFS	SOGFS	RNE	UDFS	SPCAFS	FSPCA	GSPCA	BSUFS
COIL20	76.04±1.69 (10)	69.01±1.53 (100)	<b>73.98±1.79 (80)</b>	68.03±1.59 (100)	70.76±2.07 (100)	69.12±1.17 (80)	70.29±1.31 (100)	68.41±1.60 (100)	70.44±1.37 (100)	<b>74.78±1.79 (100)</b>
Isolet	76.09±1.77 (10)	69.86±1.26 (100)	68.29±1.05 (50)	67.15±1.45 (90)	64.74±1.28 (100)	56.73±1.05 (100)	69.18±1.33 (100)	<b>71.12±1.11 (80)</b>	67.02±1.43 (100)	<b>75.32±1.22 (100)</b>
USPS	62.11±2.24 (10)	59.37±1.98 (100)	<b>62.18±2.01 (40)</b>	53.36±1.83 (100)	52.77±2.01 (100)	57.76±2.02 (100)	60.28±2.17 (100)	<b>61.14±1.87 (100)</b>	60.54±2.29 (100)	60.16±1.68 (50)
umist	64.07±1.76 (10)	61.23±2.15 (100)	<b>71.71±2.29 (100)</b>	61.46±2.03 (70)	56.08±1.80 (60)	55.43±1.50 (80)	66.26±1.74 (100)	64.94±1.65 (100)	55.88±1.62 (100)	<b>67.62±1.91 (70)</b>
GLIOMA	49.59±6.76 (10)	<b>48.96±3.59 (100)</b>	34.15±9.10 (50)	46.51±9.11 (20)	<b>54.21±2.23 (100)</b>	45.86±8.08 (20)	22.01±4.88 (80)	22.17±5.17 (90)	21.09±4.65 (100)	45.14±8.66 (100)
pie	51.01±1.02 (10)	57.53±0.73 (90)	57.16±1.01 (70)	48.05±0.76 (100)	40.45±0.79 (100)	50.55±1.03 (100)	64.94±1.30 (100)	56.21±0.90 (100)	<b>65.20±1.42 (100)</b>	<b>66.66±1.14 (80)</b>
LUNG	63.18±5.48 (10)	57.44±6.44 (70)	68.53±5.20 (100)	63.62±5.41 (40)	63.74±5.30 (90)	64.27±5.35 (90)	67.91±6.23 (100)	62.23±4.80 (20)	<b>68.96±5.71 (100)</b>	<b>72.64±4.69 (90)</b>
MSTAR	83.96±3.14 (10)	73.90±1.62 (100)	<b>81.85±2.91 (100)</b>	76.56±1.54 (100)	78.26±2.51 (100)	78.18±3.64 (90)	79.62±2.30 (100)	78.87±2.52 (90)	80.53±2.41 (100)	<b>80.66±2.68 (100)</b>
Average	65.76±2.98	62.16±2.41	<b>64.73±3.17</b>	60.59±2.96	60.13±2.25	59.74±2.98	62.56±2.66	60.64±2.45	61.21±2.61	<b>67.87±2.97</b>

# Effects of $p$ and $q$

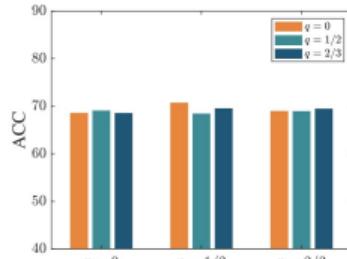
## ► ACC comparisons



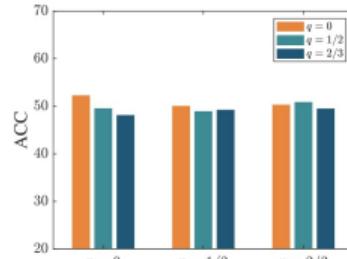
(a) COIL20



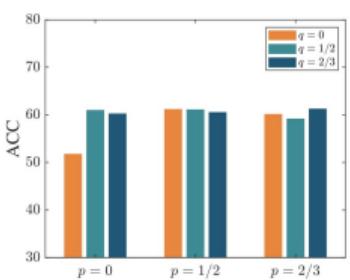
(b) Isolet



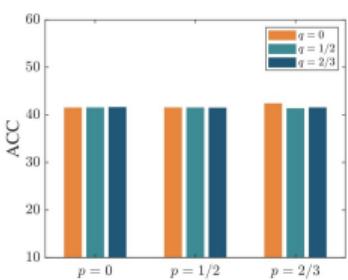
(c) USPS



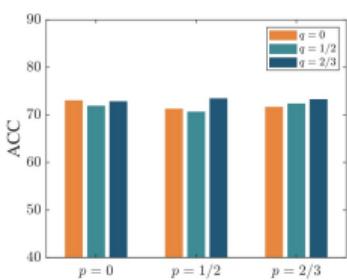
(d) umist



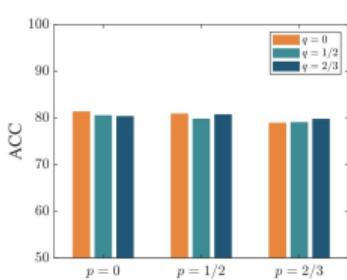
(e) GLIOMA



(f) pie



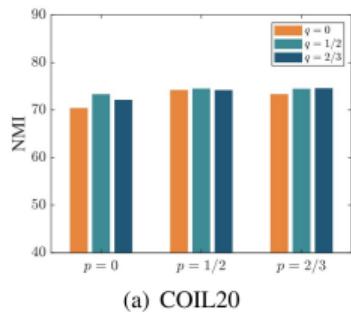
(g) LUNG



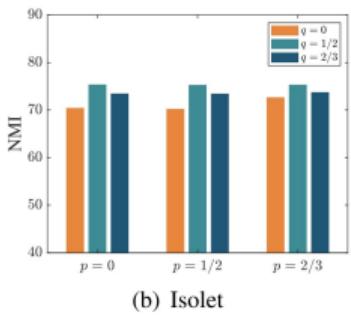
(h) MSTAR

# Effects of $p$ and $q$

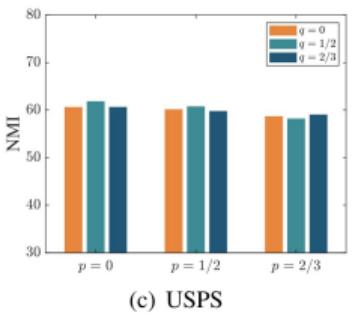
## ► NMI comparisons



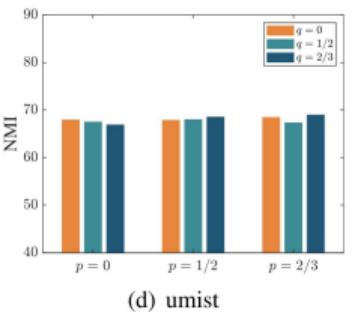
(a) COIL20



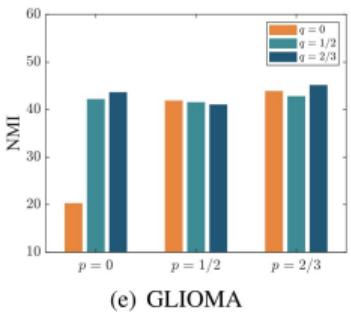
(b) Isolet



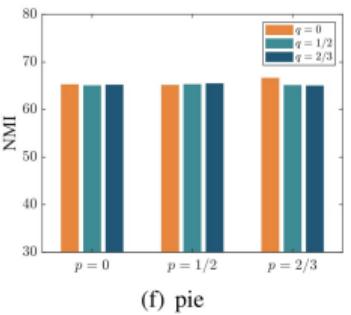
(c) USPS



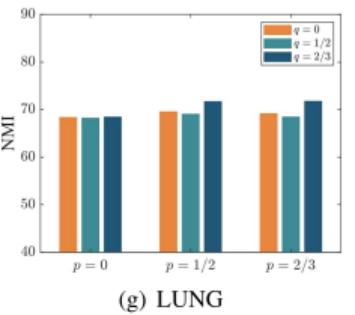
(d) umist



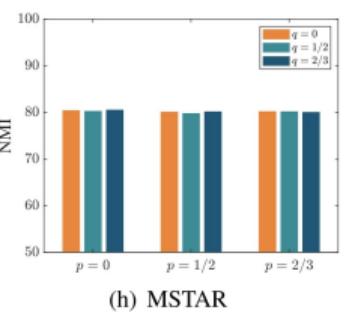
(e) GLIOMA



(f) pie



(g) LUNG



(h) MSTAR

# Ablation Experiments

## ► ACC comparisons

Datasets	Case I	Case II	Case III	Case IV
<b>COIL20</b>	54.09	57.19	<b>58.76</b>	<b>59.18</b>
<b>Isolet</b>	51.77	<b>59.49</b>	56.19	<b>61.34</b>
<b>USPS</b>	67.06	67.58	<b>68.11</b>	<b>70.77</b>
<b>umist</b>	47.16	47.52	<b>49.23</b>	<b>52.29</b>
<b>GLIOMA</b>	49.76	59.16	<b>60.12</b>	<b>61.28</b>
<b>pie</b>	40.98	40.79	<b>41.15</b>	<b>42.45</b>
<b>LUNG</b>	71.34	70.30	<b>72.33</b>	<b>73.51</b>
<b>MSTAR</b>	79.25	78.63	<b>80.08</b>	<b>81.43</b>

## ► NMI comparisons

Datasets	Case I	Case II	Case III	Case IV
<b>COIL20</b>	69.94	72.31	<b>74.57</b>	<b>74.78</b>
<b>Isolet</b>	66.84	<b>73.18</b>	72.73	<b>75.32</b>
<b>USPS</b>	<b>60.65</b>	59.10	<b>61.14</b>	60.16
<b>umist</b>	66.48	67.14	<b>69.45</b>	<b>67.62</b>
<b>GLIOMA</b>	20.64	<b>44.74</b>	43.13	<b>45.14</b>
<b>pie</b>	65.02	65.13	<b>65.23</b>	<b>66.66</b>
<b>LUNG</b>	69.31	69.08	<b>71.94</b>	<b>72.64</b>
<b>MSTAR</b>	79.92	79.61	<b>79.97</b>	<b>80.66</b>

## ► Visual comparisons

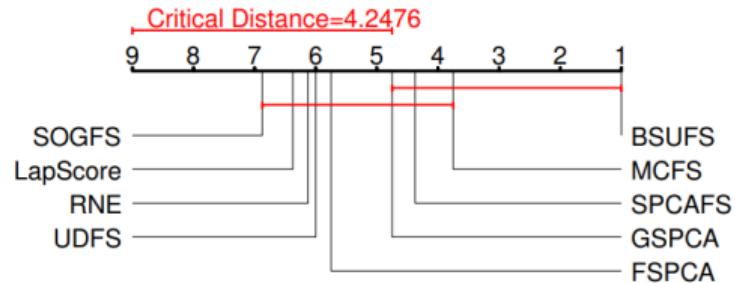
Methods	Samples				ACC	NMI
	Case I	Case II	Case III	Case IV		
Case I					40.98	65.02
Case II					40.79	65.11
Case III					<b>41.15</b>	<b>65.23</b>
Case IV					<b>42.45</b>	<b>66.66</b>

# Statistical Tests

## ► Friedman test

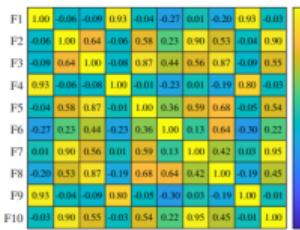
Methods	Ranking	P-value	Hypothesis
LapScore	5.670	0.0005	Reject
MCFS	<b>3.750</b>		
SOGFS	6.875		
RNE	6.125		
UDFS	6.000		
SPCAFS	4.375		
FSPCA	5.750		
GSPCA	4.750		
BSUFS	<b>1.000</b>		

## ► Post-hoc Nemenyi test

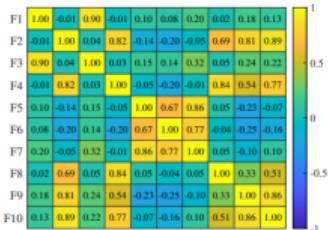


# Discussion

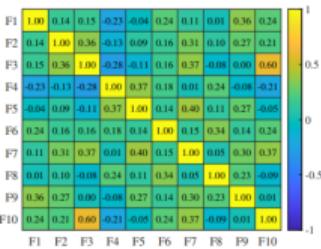
## ► Feature correlation



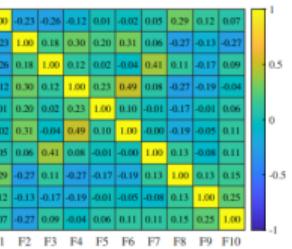
(a) COIL20 (SPCAFS)



(b) USPS (SPCAFS)

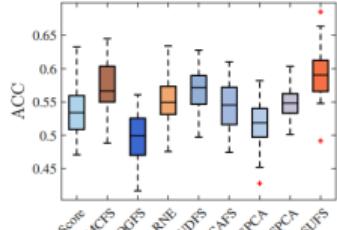


(c) COIL20 (BSUFS)

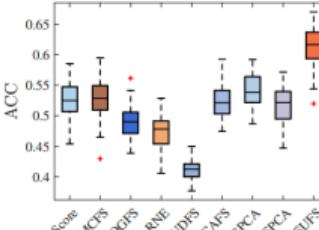


(d) USPS (BSUFS)

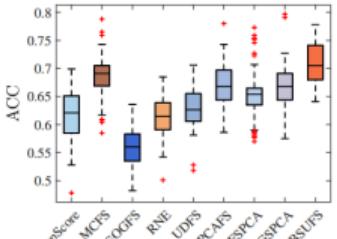
## ► Model stability



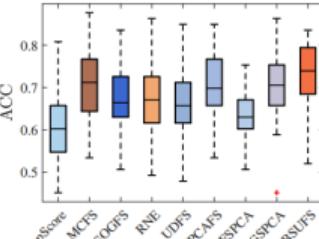
(a) COIL20



(b) Isolet



(c) USPS



(d) LUNG

# Outline

Introduction

Proposed Method

Numerical Experiments

Conclusions and Future Work

# Conclusions and Future Work

- ▶ Conclusions
  - ▶ Construct bi-sparse optimization with  $p, q \in [0, 1)$
  - ▶ Develop an efficient and convergent PAM algorithm
  - ▶ Perform sufficient experiments on real-word datasets
- ▶ Future work
  - ▶ Learn sparse via deep NNs
  - ▶ Extend to decentralized optimization
  - ▶ Apply to IoT anomaly detection

Thank you for your attention!

[xcxiu@shu.edu.cn](mailto:xcxiu@shu.edu.cn)

## References

- ▶ Y. Yang, H. Shen, Z. Ma, Z. Huang, X. Zhou,  $\ell_{2,1}$ -Norm Regularized Discriminative Feature Selection for Unsupervised Learning, IJCAI International Joint Conference on Artificial Intelligence, 2011
- ▶ Y. Zhu, X. Zhang, G. Wen, W. He, D. Cheng, Double Sparse Representation Feature Selection Algorithm for Classification, Multimedia Tools and Applications, 2017
- ▶ L. Tian, F. Nie, R. Wang, X. Li, Learning Feature Sparse Principal Subspace, Advances in Neural Information Processing Systems, 2020
- ▶ Z. Li, F. Nie, J. Bian, D. Wu, X. Li, Sparse PCA via  $\ell_{2,p}$ -Norm Regularization for Unsupervised Feature Selection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023
- ▶ G. Li, Z. Yu, K. Yang, M. Lin, C. L. P. Chen, Exploring Feature Selection With Limited Labels: A Comprehensive Survey of Semi-Supervised and Unsupervised Approaches, IEEE Transactions on Knowledge and Data Engineering, 2024
- ▶ X. Xiu, C. Huang, P. Shang, W. Liu, Bi-Sparse Unsupervised Feature Selection, IEEE Transactions on Image Processing, 2025