# Bi-Sparse Unsupervised Feature Selection

Xianchao Xiu

Department of Automation

上海大学
Shanghai University

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

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

# Outline

Introduction

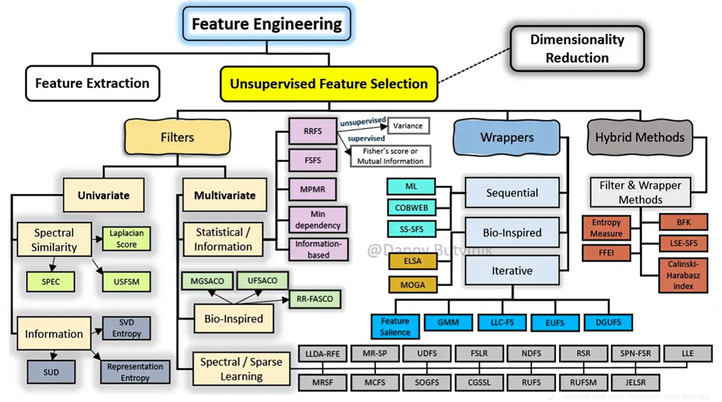
Proposed Method

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

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

- Tian-Nie-Wang-Li, NIPS, 2020

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

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

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

# Related Works

- Zhu-Zhang-Wen-He-Cheng, MTA, 2017

$$\min_{W} \; -\mathrm{Tr}(W^\top S W) + \lambda_1 \|W\|_{2,1} + \lambda_2 \|W\|_1$$
$$\text{s.t. } W^\top W = I$$

- 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

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

$$\Downarrow$$

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

- 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

$$\min_{W} \; -\mathrm{Tr}(W^\top S W) + \lambda_1 \|W\|_{2,p}^p + \lambda_2 \|W\|_q^q$$
$$\text{s.t. } W^\top W = I$$

$$\Downarrow$$

$$\min_{W,U,V} \; -\mathrm{Tr}(W^\top S W) + \lambda_1 \|V\|_{2,p}^p + \lambda_2 \|U\|_q^q$$
$$\text{s.t. } W^\top W = I, V = W, U = W$$

$$\Downarrow$$

$$\min_{W,U,V} \; -\mathrm{Tr}(W^\top S W) + \lambda_1 \|V\|_{2,p}^p + \lambda_2 \|U\|_q^q + \frac{\beta_1}{2}\|W - U\|_{\mathrm{F}}^2 + \frac{\beta_2}{2}\|W - V\|_{\mathrm{F}}^2 + \Phi(W)$$

# 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

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

  - Update $U^{k+1}$ by

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

  - Update $V^{k+1}$ by

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

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

# Update $W$

- Riemannian gradient

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

$$\Downarrow$$

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

$$\Downarrow$$

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

# Update $W$

- Riemannian Hessian

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

$$\Downarrow$$

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

$$\Downarrow$$

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

## 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

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

  - 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
    $W^k_{i+1} = R_W(\eta_i)$
  - else
    $W^k_{i+1} = W^k_i$
- Output: $W$

# Update $U$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

# Lemma

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

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

where

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

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

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

# Update $V$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

# Convergence

- $f(Q)$ is proper and lower semicontinuous.
- $f(Q)$ satisfies the K-L property at each $Q \in \operatorname{dom} f$.
- Assume the sequence $\{Q^k\}_{k\in\mathbb{N}}$ is generated by above algorithm. Then the following inequality holds
$$f(Q^{k+1}) + \tau\|Q^{k+1} - Q^k\|_{\mathrm{F}}^2 \leq f(Q^k)$$
where $\tau = \frac{1}{2}\min\{\tau_1, \tau_2, \tau_3\}$.
- Assume that $\{Q^k\}_{k\in\mathbb{N}}$ is generated by above algorithm. Then, $\{Q^k\}_{k\in\mathbb{N}}$ is bounded. In addition, there exists $V \in \partial f(Q^{k+1})$ such that
$$\|V\|_{\mathrm{F}} \leq a\|Q^{k+1} - Q^k\|_{\mathrm{F}}$$
where $a = \max\{\tau_1, \beta_1 + \tau_2, \beta_2 + \tau_3\}$.

# Convergence

- Assume $\{(W^k, U^k, V^k)\}_{k\in\mathbb{N}}$ is generated by above algorithm. Then, the sequence $\{(W^k, U^k, V^k)\}_{k\in\mathbb{N}}$ globally converges to a critical point of $f(W, U, V)$, *i.e.*,

$$0 \in \partial f(W^*, U^*, V^*)$$

  with

$$\lim_{k\to+\infty}(W^k, U^k, V^k) = (W^*, U^*, V^*)$$

  and $\partial f(\cdot)$ being the limiting subdifferential set.

# 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:

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

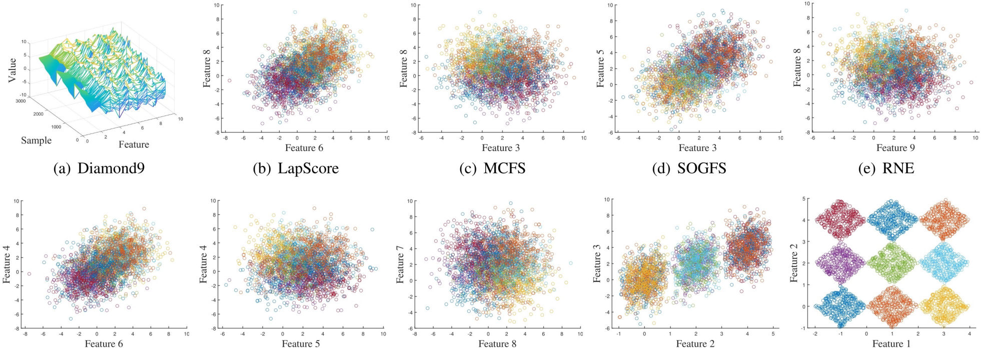
# Experimental Details

- 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

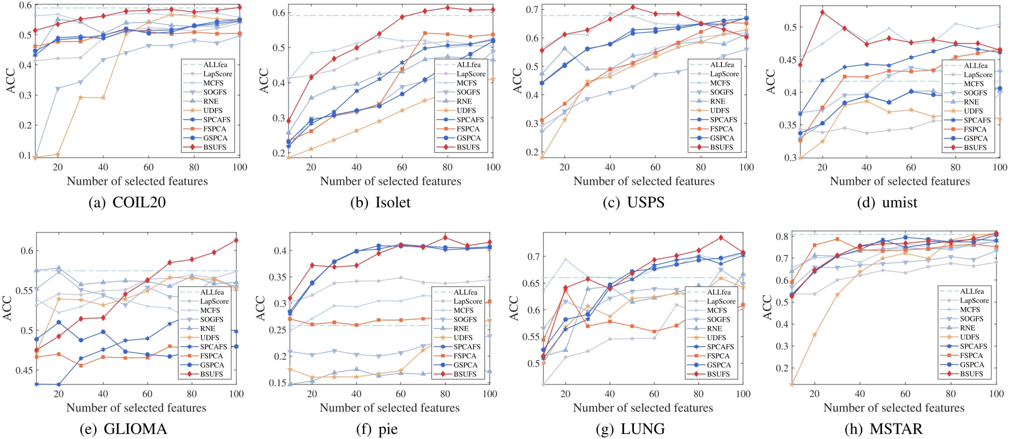



(a) Diamond9 (b) LapScore (c) MCFS (d) SOGFS (e) RNE

(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

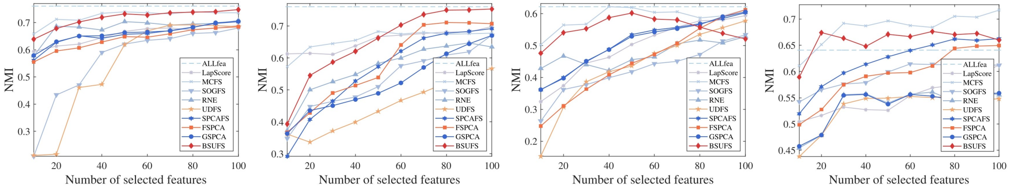
# Real Experiments

- ACC comparisons

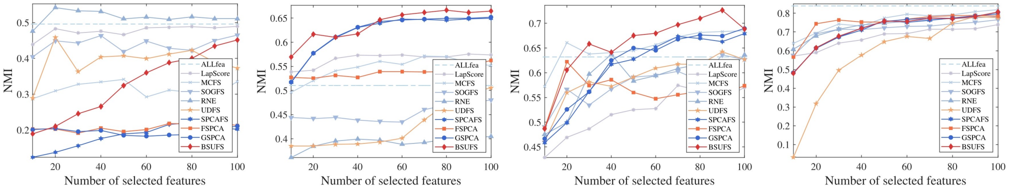
| Datasets | ALLfea | LapScore | MCFS | SOGFS | RNE | UDFS | SPCAFS | FSPCA | GSPCA | BSUFS |
|---|---|---|---|---|---|---|---|---|---|---|
| COIL20 | 58.97±4.99 (10) | 53.91±3.61 (100) | **57.35±3.83 (80)** | 49.66±3.63 (100) | 55.16±3.35 (20) | 56.77±3.09 (70) | 54.63±3.64 (100) | 51.71±3.05 (50) | 55.12±2.67 (100) | **59.18±3.49 (100)** |
| 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) | **54.15±2.69 (70)** | 51.84±2.82 (100) | **61.34±3.33 (80)** |
| USPS | 67.79±4.96 (10) | 61.76±4.52 (100) | **68.70±4.10 (40)** | 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) | **70.77±3.73 (50)** |
| umist | 41.68±2.46 (10) | 39.71±3.28 (100) | **50.54±4.16 (80)** | 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) | **52.29±3.61 (20)** |
| GLIOMA | 57.44±6.40 (10) | 57.36±3.60 (100) | 55.52±9.25 (100) | 57.32±6.47 (20) | **57.80±2.98 (20)** | 56.64±6.47 (70) | 52.08±3.64 (80) | 48.04±5.26 (90) | 51.00±5.08 (20) | **61.28±9.01 (100)** |
| 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) | **41.16±2.46 (60)** | 30.39±1.43 (100) | 40.90±1.85 (50) | **42.45±1.74 (80)** |
| 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) | **70.68±7.41 (100)** | **73.51±6.80 (90)** |
| 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) | **81.25±7.48 (100)** | 78.63±8.68 (90) | 78.74±5.20 (30) | 80.65±6.47 (100) | **81.43±6.89 (100)** |
| Average | 57.21±4.92 | 53.66±3.98 | **58.07±5.21** | 52.56±4.22 | 53.08±3.77 | 53.74±4.11 | 57.90±4.54 | 54.83±3.79 | 57.21±4.09 | **62.78±4.83** |

# Real Experiments

- NMI comparisons




(a) COIL20 (b) Isolet (c) USPS (d) umist




(e) GLIOMA (f) pie (g) LUNG (h) MSTAR

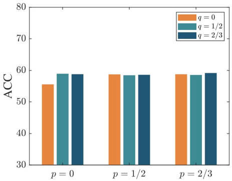
# Real Experiments

- NMI comparisons

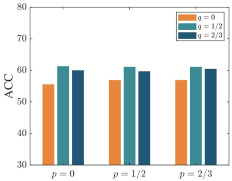
| Datasets | ALLfea | LapScore | MCFS | SOGFS | RNE | UDFS | SPCAFS | FSPCA | GSPCA | BSUFS |
|---|---|---|---|---|---|---|---|---|---|---|
| COIL20 | 76.04±1.69 (10) | 69.01±1.53 (100) | **73.98±1.79 (80)** | 68.03±1.59 (100) | 70.76±2.07 (100) | 69.12±1.17 (80) | 70.29±1.31 (100) | 68.41±1.60 (100) | 70.44±1.37 (100) | **74.78±1.79 (100)** |
| 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) | **71.12±1.11 (80)** | 67.02±1.43 (100) | **75.32±1.22 (100)** |
| USPS | 62.11±2.24 (10) | 59.37±1.98 (100) | **62.18±2.01 (40)** | 53.36±1.83 (100) | 52.77±2.01 (100) | 57.76±2.02 (100) | 60.28±2.17 (100) | **61.14±1.87 (100)** | 60.54±2.29 (100) | 60.16±1.68 (50) |
| umist | 64.07±1.76 (10) | 61.23±2.15 (100) | **71.71±2.29 (100)** | 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) | **67.62±1.91 (70)** |
| GLIOMA | 49.59±6.76 (10) | **48.96±3.59 (100)** | 34.15±9.10 (50) | 46.51±9.11 (20) | **54.21±2.23 (100)** | 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) | **65.20±1.42 (100)** | **66.66±1.14 (80)** |
| 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) | **68.96±5.71 (100)** | **72.64±4.69 (90)** |
| MSTAR | 83.96±3.14 (10) | 73.90±1.62 (100) | **81.85±2.91 (100)** | 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) | **80.66±2.68 (100)** |
| Average | 65.76±2.98 | 62.16±2.41 | **64.73±3.17** | 60.59±2.96 | 60.13±2.25 | 59.74±2.98 | 62.56±2.66 | 60.64±2.45 | 61.21±2.61 | **67.87±2.97** |

# Effects of $p$ and $q$

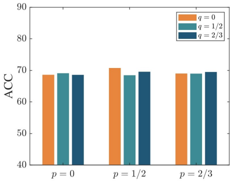
- ACC comparisons



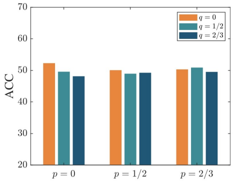

(a) COIL20



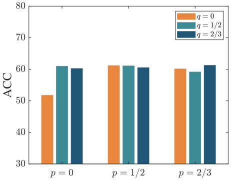

(b) Isolet



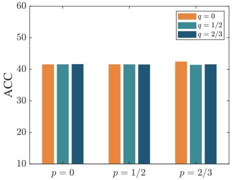

(c) USPS



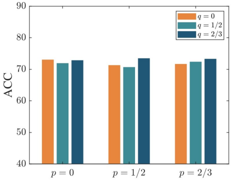

(d) umist



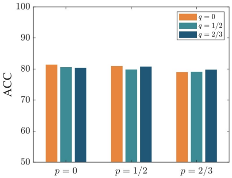

(e) GLIOMA




(f) pie



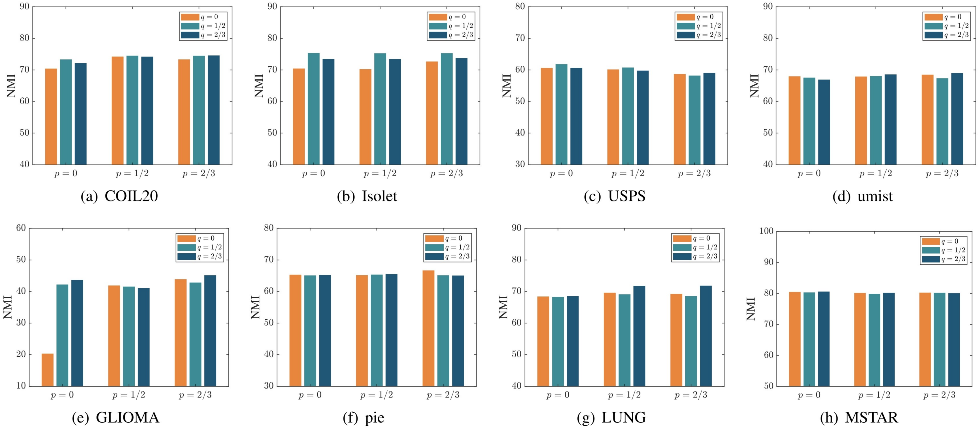

(g) LUNG




(h) MSTAR

# Effects of $p$ and $q$

- NMI comparisons



# Ablation Experiments

- ACC comparisons

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

- NMI comparisons

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

- Visual comparisons

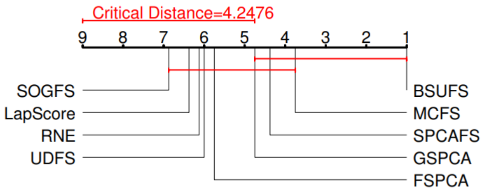
| Methods | Samples | | | | ACC | NMI |
|---|---|---|---|---|---|---|
| **Case I** |  |  |  |  | 40.98 | 65.02 |
| **Case II** |  |  |  |  | 40.79 | 65.11 |
| **Case III** |  |  |  |  | **41.15** | **65.23** |
| **Case IV** |  |  |  |  | **42.45** | **66.66** |

# Statistical Tests

- Friedman test

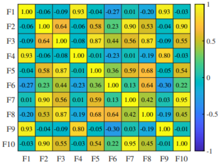
| Methods | Ranking | P-value | Hypothesis |
|---|---|---|---|
| **LapScore** | 5.670 | 0.0005 | Reject |
| **MCFS** | **3.750** | | |
| **SOGFS** | 6.875 | | |
| **RNE** | 6.125 | | |
| **UDFS** | 6.000 | | |
| **SPCAFS** | 4.375 | | |
| **FSPCA** | 5.750 | | |
| **GSPCA** | 4.750 | | |
| **BSUFS**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

- Post-hoc Nemenyi test



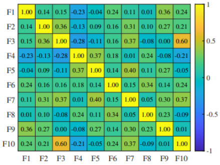
# Discussion

## ▶ Feature correlation

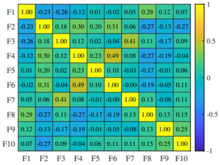



(a) COIL20 (SPCAFS)


(b) USPS (SPCAFS)

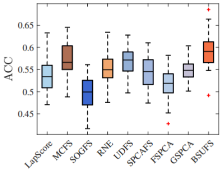



(c) COIL20 (BSUFS)

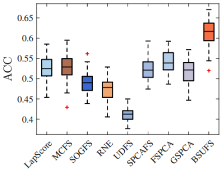



(d) USPS (BSUFS)

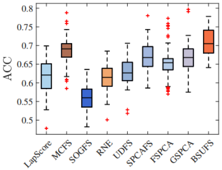
## ▶ Model stability



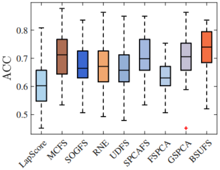

(a) COIL20




(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**

# 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, 1589-1594, 2011
- Y. Zhu, X. Zhang, G. Wen, W. He, D. Cheng, Double Sparse Representation Feature Selection Algorithm for Classification, Multimedia Tools and Applications, 76: 17525-17539, 2017
- L. Tian, F. Nie, R. Wang, X. Li, Learning Feature Sparse Principal Subspace, Advances in Neural Information Processing Systems, 33, 14997-15008, 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, 45(4): 5322-5328, 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, 36(11): 6124-6144, 2024
- X. Xiu, C. Huang, P. Shang, W. Liu, Bi-Sparse Unsupervised Feature Selection, arXiv:2412.16819