**ID:** TIP-33549-2024
**Title:** Bi-Sparse Unsupervised Feature Selection
**Authors:** Xianchao Xiu, Chenyi Huang, Pan Shang, and Wanquan Liu

## Authors' Responses

We are very grateful for your careful reading and comments on our manuscript. We have carefully read your reports and revised the manuscript according to your suggestions. Changes are in blue in the manuscript for easy check. We have addressed all the comments in detail and hope the new version is satisfactory for possible publication in *IEEE Transactions on Image Processing*.

## Responses to the Associate Editor

**Overall Comment:**
I am writing to you concerning the above referenced manuscript, which you submitted to the IEEE Transactions on Image Processing. First, I apologize as it was difficult to find reviewers. Based on the four enclosed reviews, I inform you that your manuscript needs a MAJOR REVISION (RQ).

In summary, three reviewers raised several minor concerns and corrections to be made. One reviewer raised more serious technical and positioning problems that need to be solved. As the four do not agree, I prefer to give a chance to authors to submit a revised manuscript and a point-to-point response to reviewers. Note that this revision does not mean that the revised paper will be accepted at the second round, and so the manuscript can be rejected. In the present situation, you are also free to stop the submission/revision process if you consider that you will not be able to address important concerns raised by Reviewer 3.

Your revised manuscript must be submitted back to the submission site no later than 6 weeks from the date of this letter together with a required point-by-point reply that explains how you addressed the reviewers' comments. If we do not receive your revised manuscript within 6 weeks from the date of this letter, your manuscript will be considered withdrawn.

**Response:**
Thank you for spending time and effort to coordinate the review process. In this revision, all comments and suggestions from the reviewers have been seriously taken into account and thoroughly implemented.

**If there are any further modifications that would be beneficial, we are fully prepared to implement them. We hope the revision now meets your expectation and is suitable for publication in *IEEE Transactions on Image Processing*.**

# Responses to the Reviewer #1

**Overall Comment:**
In this paper, the authors introduce a novel bi-sparse UFS (BSUFS) method, which integrates the $\ell_{2,p}$-norm and $\ell_q$-norm with $p \in [0, 1)$ into traditional principal component analysis (PCA). The proposed method can simultaneously characterize both global and local structures by selecting relevant features and filtering out irrelevant noise accurately. Moreover, the authors design a convergent proximal alternating minimization (PAM) algorithm for the BSUFS, where the orthogonality constraint is solved by trust region Riemann manifold optimization. The numerical results further validate the advantages of the proposed bi-sparse optimization in feature selection and show its potential for other fields in image processing. Overall, the paper is well written and presented. I recommend its acceptance for publication after the minor revision.

**Response:**
Thank you very much for your time and efforts devoted to the review of our paper. Your precise summary and encouraging comments are also much appreciated. All your comments or suggestions have been addressed in this revision.

**Comment 1:**
In formula (3), I think "$x_1, \cdots, x_d$" should be revised to "$\mathbf{x}^1, \cdots, \mathbf{x}^d$". The authors could examine it carefully.

**Response:**
Thanks for pointing out this mistake. According to your comment, we have revised it as follows.

(Page 3) *In details, the matrix $W$ is denoted as*

$$W = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m) = \begin{pmatrix} \mathbf{w}^1 \\ \mathbf{w}^2 \\ \vdots \\ \mathbf{w}^d \end{pmatrix} \in \mathbb{R}^{d \times m}.$$

*By transforming the vector*

$$\mathbf{x}_i = \begin{pmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{d,i} \end{pmatrix} \in \mathbb{R}^d$$

*via the transformation matrix $W$, we get the transformation vector $\mathbf{z}_i$ as*

$$\mathbf{z}_i = W^\top \mathbf{x}_i = (\mathbf{w}^{1\top},\ \mathbf{w}^{2\top},\ \ldots\ , \mathbf{w}^{d\top}) \begin{pmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{d,i} \end{pmatrix}.$$

**Comment 2:**
In Algorithm 2, it will be better if the caption “Solution of (11)” is revised to “Algorithm for solving (11)”.

**Response:**
Thanks for your careful suggestion. We have revised the title of Algorithm 2 as

(Page 4) *Trust-region Riemannian manifold optimization algorithm for solving* (12)

**Algorithm 2** Trust-region Riemannian manifold optimization algorithm for solving (12)

**Input:** Data $S \in \mathbb{R}^{d\times d}$, $U^k \in \mathbb{R}^{d\times m}$, $V^k \in \mathbb{R}^{d\times m}$, parameters $\beta_1$, $\beta_2$, $\tau_1$, $\varepsilon$, $\Delta' > 0$, $\rho' \in [0, \frac{1}{4})$
**Output:** $W^k$
**Initialize:** $i = 0$, $W_i^k \in \mathrm{St}(d, m)$, $\Delta_i \in (0, \Delta')$
**While** not converged **do**

```
Obtain M_i by solving (22) with W = W_i^k and Δ = Δ_i
Compute ρ_i from (23) with W = W_i^k
if ρ_i < 1/4 then
    Δ_{i+1} = (1/4)Δ_i
else if ρ_i > 3/4 and ‖M_i‖ = Δ_i then
    Δ_{i+1} = min(2Δ_i, Δ')
else
    Δ_{i+1} = Δ_i
end if
if ρ_i > ρ' then
    W_{i+1}^k = R_W(M_i)
else
    W_{i+1}^k = W_i^k
end if
Check convergence: If
        grad g(W_{i+1}^k) < 10^{-6} or i > 100
    then stop. Otherwise, i = i + 1
```

**End While**

**Comment 3:**
In formula (32), the optimal value of $\mathbf{v}^i$ can BE easily obtained by applying Lemma 3.1. However, it is not so obvious that the solution $\mathbf{v}^i$ is a multiplication of the normalized $\mathbf{z}^i$. It will be better if the authors could give more details and explanations.

**Response:**
Thank you for pointing out this critical issue. Formula (32) is

$$\min_{\mathbf{v}^i \in \mathbb{R}^m} \lambda_1 \|\mathbf{v}^i\|^p + \frac{\beta_2 + \tau_3}{2} \|\mathbf{v}^i - \mathbf{z}^i\|^2.$$

In this revision, we have provided more details and explanation as follows.

(Page 5) *The solution of (32) is the proximal operator of $\|\cdot\|^p$ as in [27, 44], which is reviewed in the next lemma.*
***Lemma 3.2*** *Consider the proximal operator*

$$\begin{aligned}\mathrm{Prox}_{\lambda\|\cdot\|^p}(\mathbf{a}) &= \underset{\mathbf{x}\in\mathbb{R}^m}{\mathrm{argmin}}\ \lambda\|\mathbf{x}\|^p + \frac{1}{2}\|\mathbf{x}-\mathbf{a}\|^2 \\ &= \begin{cases} \mathrm{Prox}_{\lambda|\cdot|^p}(\|\mathbf{a}\|)\cdot\frac{\mathbf{a}}{\|\mathbf{a}\|}, & \mathbf{a}\neq\mathbf{0}, \\ \{\mathbf{0}\}, & \mathbf{a}=\mathbf{0}. \end{cases}\end{aligned}$$

*Here, $\|\mathbf{x}\|^0 = 1$ when $\mathbf{x} \neq \mathbf{0}$, and $\|\mathbf{x}\|^0 = 0$ when $\mathbf{x} = \mathbf{0}$.*
*From results in Lemma 3.2, the solution of (32) is*

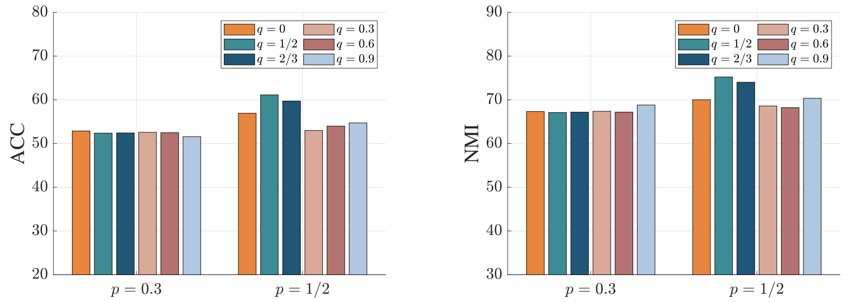
$$\mathbf{v}^i \in \begin{cases} \mathrm{Prox}_{\frac{\lambda_1}{\beta_2+\tau_3}|\cdot|^p}\left(\|\mathbf{z}^i\|\right)\cdot\frac{\mathbf{z}^i}{\|\mathbf{z}^i\|}, & \mathbf{z}^i\neq\mathbf{0}, \\ \{\mathbf{0}\}, & \mathbf{z}^i=\mathbf{0}. \end{cases}$$

[27] X. Zhang, J. Zheng, D. Wang, G. Tang, Z. Zhou, and Z. Lin, "Structured sparsity optimization with non-convex surrogates of $\ell_{2,0}$-norm: A unified algorithmic framework," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 6386–6402, 2023.
[44] A. Beck, First-order Methods in Optimization. SIAM, 2017.

**Comment 4:**
In the numerical part, the authors take the values of $p$ and $q$ for BSUFS from $0, 1/2, 2/3$ as suggested in [40] and [43]. How about the numerical performance by choosing other values of $p$ and $q$ from $[0, 1)$?

**Response:**
Thank you for pointing out this issue. We will answer it in the following two aspects.

- (**Clustering performance**) Fig. 1(a) shows the ACC results when $p = 0.3, 1/2$ and $q = 0, 1/2, 2/3, 0.3, 0.6, 0.9$ on the Isolet dataset, and Fig. 1(b) shows the corresponding NMI results. It can be seen that (1) When $p = 0.3$, different $q$ has little effect on ACC and NMI. In addition, $q = 0.9$ will lead to poor ACC but good NMI, which is not a good choice. (2) When $p = 1/2$, $q = 1/2, 2/3$ can obtain better ACC and NMI results than other values. Now, we can conclude that $0, 1/2, 2/3$ are representative choices in $[0, 1)$.




(a) ACC (b) NMI

Fig. 1. Effects of $p$ and $q$ on the Isolet dataset.

- (**Numerical calculation**) As described in [45, 43], when $p$ and $q$ are selected from the values in $\{0, 1/2, 2/3\}$, there exist closed-form proximal operators. However, for other values from $[0, 1)$, iterative calculations are necessary to obtain the optimal solutions. Thus, from the perspective of calculation, $\{0, 1/2, 2/3\}$ will be better than others.

Baes on the above two reasons, we choose $0, 1/2, 2/3$ in our experiments. In this revision, we have added the following comments.

(Page 5) *Based on [44], the proximal operator in (27) admits a closed-form when $q = 0$. This closed-form also exists when $q = 1/2$ and $q = 2/3$ as illustrated in [45, 46]. For other choices $q \in (0, 1)$, efficient algorithms proposed in [43, 47] can be considered.*

(Page 6) *As suggested in [45, 43], the values of $p$ and $q$ for BSUFS are selected from $\{0, 1/2, 2/3\}$. Although other values in $[0, 1)$ are also possible in principle, they are not considered in this paper because there is no closed-form proximal operator and the corresponding solution must be obtained by iterative calculations.*


[43] S. Zhou, X. Xiu, Y. Wang, and D. Peng, "Revisiting $L_q(0 \le q < 1)$ norm regularized optimization," arXiv:2306.14394, 2023.
[44] A. Beck, First-order Methods in Optimization. SIAM, 2017.
[45] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 7, pp. 1013–1027, 2012.
[46] W. Cao, J. Sun, and Z. Xu, "Fast image deconvolution using closed-form thresholding formulas of $L_q(q = \frac{1}{2}, \frac{2}{3})$ regularization," Journal of Visual Communication and Image Representation, vol. 24, no. 1, pp. 31–41, 2013.
[47] J. Liu, M. Feng, X. Xiu, W. Liu, and X. Zeng, "Efficient and robust sparse linear discriminant analysis for data classification," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 9, no. 1, pp. 617–629, 2025.

**Comment 5:**
Maybe it is better if the subsections "C. Ablation Experiments", "D. Effects of $p$ and $q$", "E. Discussion" are put before the subsection "B. Numerical Experiments", which shows the numerical comparisons of the proposed BSUFS method and the other ones.

**Response:**
Thank you for your suggestion. We fully appreciate your structural arrangement of the experimental part. Since "C. Ablation Experiments", "D. Effects of $p$ and $q$", "E. Discussion" are relatively long, if we put them before "B. Numerical Experiments", it will affect the reader's intuitive understanding of the experimental results.

In this revision, to make this part more clear, we have added some necessary explanations as follows.

(Page 6) *To verify whether BSUFS is better than these UFS methods and test every component of BSUFS, this section is organized as follows. Subsection IV-A describes the datasets, parameter settings, and evaluation metrics. Subsection IV-B and Subsection IV-C present numerical experiments on synthetic and real-world datasets, respectively. Subsection IV-D provides ablation experiments. Subsection IV-E shows the statistical test results. Subsection IV-F analyzes the effects of $p$ and $q$. Subsection IV-G gives more discussions.*

**Overall, all your comments and suggestions have been addressed. We hope the revision now meets your expectation and is suitable for publication in *IEEE Transactions on Image Processing.***

# Responses to the Reviewer #2

**Overall Comment:**
TBD

**Response:**
TBD

**Comment 1:**
TBD

**Response:**
TBD

**Comment 2:**
TBD

**Response:**
TBD

**Comment 3:**
TBD

**Response:**
TBD

# Responses to the Reviewer #3

**Overall Comment:**
TBD

**Response:**
TBD

**Comment 1:**
TBD

**Response:**
TBD

**Comment 2:**
TBD

**Response:**
TBD

**Comment 3:**
TBD

**Response:**
TBD