Research Article

Dimensionality Reduction with Sparse Locality for Principal Component Analysis

Pei Heng Li , Taeho Lee, and Hee Yong Youn

Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea

Correspondence should be addressed to Hee Yong Youn; youn7147@skku.edu

Received 26 December 2019; Revised 6 March 2020; Accepted 6 April 2020; Published 20 May 2020

Academic Editor: Giuseppe D'Aniello

Copyright © 2020 Pei Heng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Various dimensionality reduction (DR) schemes have been developed for projecting high-dimensional data into low-dimensional representation. The existing schemes usually preserve either only the global structure or local structure of the original data, but not both. To resolve this issue, a scheme called sparse locality for principal component analysis (SLPCA) is proposed. In order to effectively consider the trade-off between the complexity and efficiency, a robust $L_{2,p}$ -norm-based principal component analysis (R2P-PCA) is introduced for global DR, while sparse representation-based locality preserving projection (SR-LPP) is used for local DR. Sparse representation is also employed to construct the weighted matrix of the samples. Being parameter-free, this allows the construction of an intrinsic graph more robust against the noise. In addition, simultaneous learning of projection matrix and sparse similarity matrix is possible. Experimental results demonstrate that the proposed scheme consistently outperforms the existing schemes in terms of clustering accuracy and data reconstruction error.

1. Introduction

In recent years, the development of high-throughput data processing schemes in diverse fields including pattern recognition, data mining, and computer vision has resulted in an exponential growth in the amount of harvested data with respect to both the dimensionality and size. However, large amount of redundancy and noise in the data cause significant spatial instability, computational complexity, and unfavorable representation. In relation to this problem, dimensionality reduction (DR) has been identified as an effective approach due to its capacity in dealing with a large amount of data and potential for overcoming what is called the “curse of dimensionality” [1]. It also offers a greater scope of model generalization and accomplishes the tasks with a high degree of computational efficiency.

To date, a variety of DR techniques have been developed for projecting the original data of high-dimensional space into a lower dimensional subspace. They can be classified in two ways: global dimensionality reduction (GDR) and local dimensionality reduction (LDR) [2]. The GDR techniques assume that all pairwise distances of the data are of equal

importance. The globally correlated data are reduced using the magnitude or rank-over to choose the optimal low-dimensional pairwise distance, thereby eliminating irrelevance and redundancy [3]. With the LDR technique, only the local distances are assumed to be reliable in high-dimensional space. More emphasis is therefore put on correctly modeling the locally correlated data to eliminate the noise [4]. A great deal of research has been conducted using realistic nonimage and image datasets, with the results showing that a significant percentage of redundancy and noise is eliminated by both the techniques [5–7].

Recently, a number of hybrid global-local methods have been proposed. The authors of [5], for instance, have developed an algorithm of hybrid sampling-based clustering ensemble with global constitution, which encodes the local and global cluster structure of input partitions in a single representation. Here deciding a final consensus candidate involves significant computational cost. In [6], a DR algorithm was proposed which uses pairwise similarity measurement to effectively capture the local structure of data manifolds. While offering considerable advantages, the hyperparameters of the selected model remain as an issue. In

[7], a global-local structure preservation framework was introduced for feature selection based on three algorithms: local linear embedding (LLE), local tangent space alignment (LTSA), and locality preserving projection (LPP). In this framework, the copious amount of noise generated during the process may degrade the reliability of the data. A well-designed DR model still needs to be developed that can effectively reduce the data redundancy and noise while ensuring robustness.

In this paper, thus, a global-local DR model called sparse locality for PCA (SLPCA) is proposed, which introduces robustness into both GDR and LDR by employing a regularizer and constraint, separately. The SLPCA model aims to reduce unnecessary information while preserving the data correlation in both the global and local structure. In order to effectively consider the trade-off between the complexity and efficiency, a robust $L_{2,p}$ -norm constraint-based PCA (R2P-PCA) is introduced for GDR, while sparse representation (SR) and LPP are combined for LDR. The R2P-PCA algorithm implements PCA with a robust distance metric $L_{2,p}$ -norm by maximizing the sum of the variations. In this way, R2P-PCA can increase the robustness against the noise. It is also for preserving the global structure of the samples. The SR-LPP algorithm seeks a set of projection matrices by capitalizing the merits of SR and LPP, which are merged into one analytical process. SR enables adaptive construction of the graph because it is parameter-free and robust against the noise [8]. This makes SR-LPP capable of simultaneous and adaptive learning of the projection matrix and sparse similarity matrix. Through the joint learning of the two algorithms, the learned functions can be optimized using an efficient iterative algorithm, alternating direction method of multipliers (ADMM) [9]. It monotonically decreases the value of the augmented Lagrangian function as the iteration continues. Computer simulation reveals that the proposed SLPCA scheme consistently outperforms the existing schemes in terms of clustering accuracy and data reconstruction error. The main contributions of the paper are summarized below:

- (i) A means of delivering robustness in DR is developed that simultaneously considers both the local and global structure of sample data.
- (ii) A scheme effectively considering the trade-off between the complexity and efficiency of DR is proposed, where a robust $L_{2,p}$ -norm constraint-based PCA (R2P-PCA) is introduced for global DR, while sparse representation (SR) and LPP are combined for local DR.
- (iii) Typically, the graph-based DR methods calculate a projection matrix on the basis of a learned graph. A novel approach is developed that can simultaneously integrate both the projection matrix and nonparametric graph construction.

The remainder of the paper is organized as follows. In Section 2, the work related to the preservation of global and local data structure is surveyed and summarized. The proposed sparse locality for PCA (SLPCA) model is introduced

in Section 3. This section also presents the key algorithms of SLPCA, R2P-PCA, and SR-LPP and explains how they achieve convergence. In Section 4, the performance of SLPCA is evaluated in terms of clustering accuracy and data reconstruction error with two kinds of datasets. The paper is concluded in Section 5, with future research directions.

2. Related Work

2.1. Notation and Definition. Given the vector $v \in R_m$, the L_p -norm of v is defined as $\|v\|_p = (\sum_{i=1}^m |v_i|^p)^{1/p}$. When $p = 0$, the L_0 -norm of v is defined as $\|v\|_0 = \sum_{i=1}^m |v_i|^0$. When $p = 1$, the L_1 -norm of vector v is $\|v\|_1 = \sum_{i=1}^m |v_i|$. Given the matrix $M_{ij} \in R^{n \times m}$, the i -th row and j -th column are denoted by m^i and m_j , respectively. L_0 -norm $\|M\|_0$ denotes the number of nonzero elements in the matrix M . The L_1 -norm of M is defined as

$$\|M\|_1 = \sum_{i=1}^n \sum_{j=1}^m |m_{ij}| = \sum_{i=1}^n \|m^i\|_1. \quad (1)$$

The Frobenius norm of M is

$$\|M\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m m_{ij}^2} = \sqrt{\sum_{i=1}^n \|m^i\|_2^2}. \quad (2)$$

The $L_{2,1}$ -norm of M was first introduced in [10] as a rotational invariant for the rows. It is now widely employed to encourage row sparsity and is defined as

$$\|M\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m m_{ij}^2} = \sum_{i=1}^n \|m^i\|_2. \quad (3)$$

2.2. DR Based on Global Structure. There exist three main approaches for the DR based on the global structure in the literature.

2.2.1. Principal Component Analysis. PCA is the most commonly used algorithm in this category because of the simplicity and efficiency. It aims to reduce the dimension of the original data by projecting them across the direction of maximum variance. The Graph-Laplacian PCA algorithm was developed by Jiang et al. [11], which learns a low-dimensional representation of vector data by incorporating the graph structure encoded in the graph data. In [12], rotational invariance L_1 -norm PCA minimizes the reconstruction error by imposing an L_2 -norm on the spatial distance, whereas an L_1 -norm is applied to different data points. The authors of [13] have proposed the convex sparse PCA approach that reduces redundant information by building a compact and informative subspace. These kinds of algorithms are useful in the fields including image processing, speech enhancement, and reduction of data transmissions.

2.2.2. Distance Preserving Method. The typical distance preserving algorithms are based on maximum variance unfolding (MVU) and isometric mapping (Isomap). Although both the approaches result in a similar computational structure, Isomap merely preserves the global geodesic distance, whereas MVU preserves the local distance but maximizes the global variance. In [14], a hybrid incremental landmark MVU model was developed that was combined with a dual-tree complex wavelet transform. Saxena et al. [15] introduced a nonlinear MVU-PCA algorithm, which aims to enhance the core components of the framework for higher robustness. The multilevel MVU technique proposed in [16] aims to reduce and distribute the computational load across parallel multilevel machines. The algorithms mentioned above share a weakness of erroneous graph connection.

2.2.3. Autoencoder. Autoencoder (AE) [17] focuses on learning the underlying manifold in the encoding and decoding procedure. The multilayer AE [18] usually has a higher number of connections, but it may cause slow convergence via backpropagation even to local minima. The spectral AE [19] aims to detect global and community anomaly by mapping the attributed network to two types of low-dimensional representations. The AE-based approaches are rather limited in their capacity in learning the local structure, which contains important information required for understanding the data.

2.3. DR Based on Local Structure. The DR schemes based on local data structure usually adopt the graph embedding technique such as LPP, LLE [20], Laplacian eigenmaps [21], LTSA [7], and neighborhood preserving embedding (NPE) [22]. They approximate the embedded manifold through a mapping function with the proximity preserving property. The LPP algorithm uses the k -NN-based graph Laplacian regularizer to preserve the local structure of the samples. Yu et al. [23] improved the LPP scheme where the L_1 -norm is used to provide the robustness while the similarity between the pairs of vertices is effectively preserved simultaneously. In [24], a supervised global-locality preserving projection approach is proposed that preserves not only the locality characteristics of LPP but also the global discriminative structure associated with maximum margin criterion. The aforementioned algorithms can be hampered by the presence of noise. The low-rank learning method can help to reduce the disturbance caused by noise in the data.

In [25], a neighborhood preserving projection method was developed that integrates low-rank learning and robust learning. It enhances the robustness of the NPP method and diminishes the disturbance caused by noise in the data. Similarly, the low-rank preserving projection method introduced in [26] applies an L_{21} norm to the noise matrix as a sparse constraint and a nuclear norm to the weight matrix as a low-rank constraint. It maintains the global structure of the data during DR, and the learned low-rank weight matrix

reduces the noise-related disturbance. The scheme of [27] preserves the structure of nonlinear manifold data, and it is robust because the learned data are unaffected by noises or outliers. In [28], a structurally incoherent low-rank non-negative matrix factorization method was proposed that jointly considers the structural incoherence and low-rank property of data. Since the scheme employs low-rank learning method, it allows to capture the global structure of the data and robustness to the noise. The proposed scheme is presented next.

3. The Proposed Scheme

3.1. R2P-PCA for Preserving Global Structure. R2P-PCA is for the preservation of the global structure of the data. PCA finds a few orthogonal directions in the data space that preserve the most information in the data while minimizing the reconstruction error. Assuming an input data matrix, $X=(x_1, \dots, x_n) \in R^{m \times n}$, that contains m data column vectors in n dimensional space, PCA finds the optimal low-dimensional subspace of U and V by solving the following optimization problem:

$$\begin{aligned} \min_{U,V} \|X - UV^T\|_F^2, \\ \text{s.t. } V^T V = I, \end{aligned} \quad (4)$$

where each column of $U=(u_1, \dots, u_k) \in R^{m \times k}$ is a principal direction. $V^T=(v_1, \dots, v_k)^T \in R^{k \times n}$ represents the projected data points in the new reduced subspace.

Recently, various PCA-based methods have employed different criterion functions such as L_1 -norm to improve the robustness to noise [29]. Although L_1 -norm is robust to outliers, most existing L_1 -norm-based PCA methods do not effectively minimize the reconstruction errors which is one of the main goals of PCA. They are also not invariant under rotation, which is an important property for learning algorithm. In the proposed scheme, we therefore focus on formulating a robust learning metric formulation, the $L_{2,p}$ -norm [30–32]. Note that most existing PCA methods based on $L_{2,1}$ -norm can be viewed as special cases of the proposed R2P-PCA approach. It is well known that $\|\cdot\|_{2,1}$ is convex with respect to the matrix variables, and thus it can be extended to a generalized robust learning metric formulation for matrix M , namely, the $L_{2,p}$ -norm ($0 < p < 2$), which can be defined as follows:

$$\|M\|_{2,p} = \sqrt[p]{\left(\sum_{i=1}^n \|m^i\|_2^p \right)}. \quad (5)$$

On the basis of equations (4) and (5), the objective function of R2P-PCA can now be obtained:

$$\begin{aligned} \min_{U,V} \|X - UV^T\|_{2,p}^p, \\ \text{s.t. } V^T V = I. \end{aligned} \quad (6)$$

3.2. SR-LPP for Preserving Local Structure

3.2.1. Original Locality Preserving Projection. LPP can be viewed as a linear approximation of nonlinear Laplacian eigenmaps. The first step in the original LPP is to construct an adjacency graph, which significantly affects the performance of the algorithm. Then, the projection matrix is calculated using the learned graph. The objective function of LPP can be formally stated as follows:

$$\min_w \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{ij}, \quad \text{s.t.} \quad Y^T Y = I. \quad (7)$$

Let p denote the transformation vector. Through simple algebraic reformulation, the objective function above can be rewritten as follows:

$$\begin{aligned} & \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{ij} \\ &= \frac{1}{2} \sum_{ij} (p^T x_i - p^T x_j)^2 W_{ij} \\ &= \sum_{ij} p^T x_i \cdot W_{ii} \cdot x_i^T p - \sum_{ij} p^T x_i \cdot W_{ij} \cdot x_j^T p \\ &= \sum_i p^T x_i D_{ii} x_i^T p - \sum_{ij} p^T x_i W_{ij} x_j^T p \\ &= P^T X (D - W) X^T P \\ &= P^T X L X^T P, \end{aligned} \quad (8)$$

where $p^T x_i$, $p_j^T v$, and W_{ij} are numerical values, $D_{ii} = \sum_j W_{ij}$, and $L = D - W$. Here, $-L$ denotes the Laplacian matrix, which is used to minimize $P^T X L X^T P$. The LPP objective function can therefore be changed to

$$\begin{aligned} & \min_P P^T X L X^T P, \\ & \text{s.t.} \quad P^T X D X^T P = I. \end{aligned} \quad (9)$$

For the heat kernel method, the weight assignment can be obtained using

$$W_{ij} = \begin{cases} \exp \left[-\frac{\|x_i - x_j\|^2}{t} \right], \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

In [33], it was established that adjacent graph structure and graph weight are highly interrelated and should not be separated. This makes it preferable to develop one ideal model that can perform the two tasks simultaneously. The traditional weight assignment methods require parameter selection to construct the adjacency graph by means of the ϵ -ball or k -NN method. This comes with a significant computational cost if they need to identify the

neighborhoods for a whole dataset. This also makes them sensitive to data noise [34]. Hence, instead of using them, here we attempt to automatically derive the similarity matrix and ensure that it preserves the discriminative information by using sparse representation (SR).

3.2.2. Graph Construction Based on SR. The main idea of SR is that a body of sampled data can be sparsely represented when given an appropriate basis. If a given test sample y , belongs to the i -th class, SR assumes that y can be represented as a linear combination of the training samples in the i -th class, $X_i = \{x_i^1, x_i^2, \dots, x_i^{N_i}\}$. In other words, y can be represented as $y = w_i^1 x_i^1 + w_i^2 x_i^2 + \dots + w_i^{N_i} x_i^{N_i} = X_i W_i$, where W_i denotes the coefficient of y over X_i . Ideally, the coefficients of other classes are zero, such that $W_j = 0$ and $\forall j \neq i$. Thus, y can be represented as a linear combination of all the training samples by stacking X_i ($i = 1, 2, \dots, C$) within the whole training set, X . Then, the coefficient matrix, $W = [W_1, W_2, \dots, W_C]$, can be obtained. So,

$$y = 0 + \dots + 0 + w_i^1 x_i^1 + w_i^2 x_i^2 + \dots + w_i^{N_i} x_i^{N_i} + 0 + \dots + 0 = XW. \quad (11)$$

The above model can be expressed as follows:

$$\begin{aligned} & \tilde{W} = \arg \min \|W\|_0, \\ & \text{s.t.} \quad y = XW. \end{aligned} \quad (12)$$

Note that the L_0 -norm optimization problem is an NP-hard problem. A recent study has demonstrated that L_0 -norm is equivalent to the L_1 -norm optimization problem if the solution is sparse enough [35], in which case it can be solved using

$$\tilde{W} = \arg \min \|W\|_1, \quad \text{s.t.} \quad y = XW. \quad (13)$$

SR avoids parameter selection and makes the intrinsic graph construction more robust to data noise. Having established the ground of theoretical inspiration related to LPP and SR, the basic idea of the proposed SR-LPP method for local structure preservation is presented next.

3.2.3. SR-LPP. SR-LPP learns the similarity matrix, W , and projection matrix, P , simultaneously and adaptively so that the intrinsic properties of the structure can be preserved. For this, the LPP objective function needs to be changed to

$$\begin{aligned} & \min_w \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \|P^T x_i - P^T x_j\|_2^2 W_{ij}, \\ & \text{s.t.} \quad P^T X D X^T P = I. \end{aligned} \quad (14)$$

In addition, in order to ensure that each sample is represented by unique base, an L_1 -norm constraint is added to W . According to equation (9) in [36], the objective function for SR-LPP can now be formulated as follows:

$$\begin{aligned}
& \min_w \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \|P^T x_i - P^T x_j\|_2^2 W_{ij} \\
& + \lambda_1 \|X - XW\|_F^2 + \lambda_2 \|W\|_1, \\
& \text{s.t. } P^T X D X^T P = I,
\end{aligned} \tag{15}$$

where λ_1 and λ_2 (≥ 0) are parameters balancing the contribution from the two parts.

3.3. The Objective Function. For joint learning with R2P-PCA preserving global structure and SR-LPP preserving local structure, equations (6) and (15) are combined in the following model:

$$\begin{aligned}
F(U, V, W) &= \min_{U, V} \|X - UV^T\|_{2,p}^p + \alpha \text{Tr}(P^T X L X^T P) \\
& + \lambda_1 \|X - XW\|_F^2 + \lambda_2 \|W\|_1, \\
& \text{s.t. } P^T X D X^T P = I,
\end{aligned} \tag{16}$$

where α (≥ 0) is a parameter balancing the contribution from the three parts. It should be pointed out that the update of P in equation (16) can be derived by solving the generalized eigenvector problem, which corresponds to the eigenvectors of the k smallest eigenvalues. Let $V = P^T X$. As v_i in PCA takes exactly the same role as $p^T x_i$ in LPP, the objective function can be simplified to

$$\begin{aligned}
F(U, V, W) &= \min_{U, V} \|X - UV^T\|_{2,p}^p + \alpha \text{Tr}(V^T \mathbb{L} V) \\
& + \lambda_1 \|X - XW\|_F^2 + \lambda_2 \|W\|_1, \\
& \text{s.t. } P^T X D X^T P = I,
\end{aligned} \tag{17}$$

where $\mathbb{L} = D - W$. ADMM can be expanded to solve the subproblem. Letting $\mathbb{E} = X - UV^T$, the standard Lagrange function in the equation above can be augmented as follows:

$$\begin{aligned}
\mathcal{L}_\rho(U, V, W, \mathbb{E}, \rho) &= \|\mathbb{E}\|_{2,p}^p + \alpha \text{tr}(V^T \mathbb{L} V) \\
& + \lambda_1 \|X - XW\|_F^2 + \lambda_2 \|W\|_1 \\
& + \left(\frac{\rho}{2}\right) \|\mathbb{E} - X + UV^T + u\|_F^2.
\end{aligned} \tag{18}$$

Hence, the iterative computation consists of W , U, V , and $\mathbb{E}$ -minimization step and an update of the Lagrangian parameter, ρ . The detailed iterative computational steps are formulated below.

3.4. Optimization Analysis

3.4.1. Fixing $\mathbb{E}$ and W to Update U and V . According to Theorem 3.1 in [11], let $\mathbb{F} = \mathbb{E} - X + u$; U can be updated by solving the following optimization problem:

$$\arg \min_U \left(\frac{\rho}{2}\right) \|\mathbb{F} + UV^T\|_F^2. \tag{19}$$

Taking the partial derivative of U and setting it to 0, the following can be obtained:

$$\frac{(\rho/2) \|\mathbb{F} + UV^T\|_F^2}{U} = 0 \Rightarrow \hat{U} = -\mathbb{F}V. \tag{20}$$

Then, U is updated. As $\hat{U} = -\mathbb{F}V$, the optimization problem with respect to V in equation (18) can be simplified into the following equation:

$$\arg \min_V \beta \text{tr}(V^T \mathbb{L} V) + \left(\frac{\rho}{2}\right) \|\mathbb{F} - \mathbb{F}V V^T\|_F^2. \tag{21}$$

With similar operation on U ,

$$\text{tr} V^T \left(-\mathbb{F}^T \mathbb{F} + \frac{2\beta}{\rho} \mathbb{L} \right) V = 0. \tag{22}$$

The optimal V^* can be obtained by calculating the eigenvectors, $\hat{V} = (v_1, \dots, v_k)^T$. This is the first k smallest eigenvalue of the matrix $\mathcal{Q} = -\mathbb{F}^T \mathbb{F} + 2\beta \mathbb{L} / \rho$. According to Proposition 3.2 in [11], the solution of V can be expressed by the eigenvectors of

$$Q_\alpha = (1 - \theta) \left(I - \frac{\mathbb{F}^T \mathbb{F}}{\sigma_m} \right) + \frac{2\theta}{\rho} \left(\frac{\mathbb{L}}{\sigma_l} + \frac{ee^T}{n} \right), \tag{23}$$

where σ_m and σ_l are the largest eigenvalue of matrix $M^T M$ and L , respectively; θ is the parameter substituting α ; and $e = (1 \dots 1)^T$ is an eigenvector of Q_α : $Q_\alpha e = (1 - \theta)e$. Although (ee^T/n) is applied in place of the Laplacian matrix, the eigenvectors and eigenvalues do not change. This guarantees that e is not included in the lowest k eigenvectors.

3.4.2. Fixing U , V , and W to update $\mathbb{E}$. Let $\mathbb{S} = X - \hat{U}(\hat{V})^T - u$. $\mathbb{E}$ can now be updated by solving the following optimization problem:

$$\arg \min_{\mathbb{E}} \|\mathbb{E}\|_{2,p}^p + \left(\frac{\rho}{2}\right) \|\mathbb{E} - \mathbb{S}\|_F^2. \tag{24}$$

Taking the partial derivative of $\mathbb{E}$ and setting it to 0, according to the proximal operator of $L_{2,p}$ norm proposed in [30], change $L_p(\hat{\mathbb{E}})$ to $\sum_{i=1}^n (\|\mathbb{E}_i\|_2^2 + \varepsilon_0)^{p/2} + (\rho/2) \|\mathbb{E} - \mathbb{S}\|_F^2$, where $\mathbb{E}_i$ is the i -th column of $\mathbb{E}$ and $0 < \varepsilon_0 \leq 1$. Then, $\mathbb{E}$ can be obtained using

$$\mathbb{E} = \rho \mathbb{S} / (pN + \rho I), \tag{25}$$

where N is the weight matrix corresponding to $\mathbb{R}(\mathbb{E})$; here, $N_{ii} = (\|\mathbb{E}_i\|_2^2 + \varepsilon_0)^{p/2-1}$. The $L_{2,p}$ -norm minimization problem can then be solved by updating N and $\mathbb{E}$ iteratively. The procedure for this is presented in Algorithm 1.

3.4.3. Fixing U , V , and $\mathbb{E}$ to Update W . W can be updated by solving the following optimization problem:

$$\arg \min_W \text{Tr}(P^T W) + \lambda_1 \|X - XW\|_F^2 + \lambda_2 \|W\|_1. \tag{26}$$

Taking the partial derivative of W and setting it to 0, apply the nonnegative projection proposed in [37]. Then, obtain W^* using

```

Input: matrix  $\mathbb{S}$ ;
Step 1: Select parameters:  $p, \rho, \varepsilon$ ;
Step 2: Optimization
Initialize:  $\mathbb{E} = 0, N0 = 0, k = 0, \varepsilon = 0.8$ .
While no convergence do
  (1) Update N using  $N_{ii} = (\|\mathbb{E}_i\|_2^2 + \varepsilon_0)^{p/2-1}$ 
  (2) Update  $\mathbb{E}$  using equation (25);
  (3) Update  $k = k + 1$ ;
  (4) Check the convergence condition:
 $\|\mathbb{E}_k - \mathbb{E}_{k-1}\| < \varepsilon$ ;
End and
Output:  $k$ -dimensional vector  $\mathbb{E}$ .

```

ALGORITHM 1: The procedure for solving equation (25).

$$\begin{cases} W^* = (2\lambda_1 X^T X)^{-1} (2\lambda_1 X^T X - \lambda_1 - P), \\ W^* = \max(0, W^*). \end{cases} \quad (27)$$

3.4.4. *Updating the Parameters.* Drawing upon [9], the parameters need to be updated as follows:

$$\begin{aligned} \hat{u} &= u + \hat{\mathbb{E}} - X + \hat{U}(\hat{V})^T, \\ \hat{y} &= y + \rho(\hat{\mathbb{E}} - X + \hat{U}(\hat{V})^T), \\ \hat{\rho} &= \gamma\rho^k. \end{aligned} \quad (28)$$

3.5. *Convergence Property.* The proposed sparse locality-regularized PCA model (SLPCA) consists of two main parts: solving the subproblems and updating the parameters. For the update of the first subproblem, the optimal solution of each column of $\mathbb{E}$ is obtained using equation (24). Therefore, the value of the objective function decreases after solving $\mathbb{E}$ in each iteration. For the other two subproblems, the minimization of the augmented Lagrangian occurs with respect to the variables, U, V , and W . When updating U , because the Lagrange function of equation (18) is also convex and differentiable to U , the optimal projection matrix can be obtained by using a closed-form solution of equation (20). This means that the value of the objective function decreases after solving U in each iteration. The same procedure is applied for updating V and W . The proposed SLPCA algorithm is therefore convergent.

The procedure for the proposed SLPCA is shown in Algorithm 2, which is applied when the subproblems are solved. It searches for an optimal value in each iteration and finally stops when the value of the objective function becomes stable.

3.6. *Computational Complexity.* With the proposed SLPCA algorithm, the major computational burden lies in Steps 2 and 4 of Algorithm 2. Step 2 is for the eigenvalue decomposition on an $m \times m$ matrix, of which computational complexity is $O(m^3)$. Step 4 is the inverse operation of an $n \times n$ matrix of the complexity of $O(n^3)$. Therefore, the

computational complexity of Algorithm 2 is $O(t(m^3 + n^3))$, where t is the number of iterations.

4. Performance Evaluation

4.1. *Experiment Setup.* This section reports an experiment conducted to verify the effectiveness of the proposed SLPCA scheme. A comparison is also made regarding the performance in finding low-dimensional representation and classification in relation to the two baseline algorithms, PCA and LPP, and three state-of-the-art hybrid global-local DR methods, i.e., GRSDA [38], L1/2-GLPCA [39], and p -GLPCA [40]. In the experiment, the value of the parameter θ is chosen from the set $[0.1, 0.5, 0.9]$, and p is chosen from the set $[0.5, 1]$. According to the result of experiment, the proposed method is superior to the others when the parameter α is larger than 0.6. The value of λ_1, λ_2 , and ρ are empirically determined, while it was found that their values do not significantly influence the performance.

In order to evaluate the performance of the schemes on various characteristics including classification and data reconstruction accuracy, the experiment is performed using two different types of datasets: nonimage-based dataset and image-based dataset. For the nonimage data, three UCI datasets are employed: Iris, Seeds, and Soybean that are publicly available at <https://archive.ics.uci.edu/ml/datasets.php>. For the image data, three commonly adopted datasets are used, the Extended Yale-B, ORL face, and CMU PIE dataset. The K-means test is applied to assess the effectiveness of the proposed scheme. The experiment is also extended to test its robustness and generalizability for different values of p . Table 1 summarizes the main features of the DR models compared.

4.2. *Experiment on Nonimage Datasets.* This section details the results of the experiment conducted to assess the effectiveness of the schemes in terms of subspace representation and clustering accuracy. The Iris dataset is one of the well-known pattern recognition databases published in the literature. It contains three classes of 50 instances each, where each class refers to a type of iris plant. Here, one class

Input: Training set $X = \{x_1, x_2, \dots, x_N\}$, the reduced dimension k (with $k \leq D$).
 Step 1: Select parameters: $\rho, \theta, \lambda_1, \lambda_2$.
 Step 2: Alternative Optimization
Initialize $\rho = 0, \theta = 0, U = 0, V = 0$.
While no convergence do
 (1) Fix the other variables and update U using equation (20);
 (2) Fix the other variables and update V using equation (22);
 (3) Fix the other variable and update the auxiliary $\mathbb{E}$ using equation (25);
 (4) Fix the other variables and update W using equation (27);
 (5) Update the Lagrange multiplier and other parameters using equation (28);
End and
Output: k -dimensional vector $U, V, W, \mathbb{E}$.

ALGORITHM 2: SLPCA procedure.

TABLE 1: The comparison of the DR models and their properties.

	Algorithm	Model	Constraints	Parameters
1	PCA	$\min_{U,V} \ X - UV^T\ _F^2$	$V^T V = I$	No
2	LPP	$Tr(Y^T L Y)$	$Y^T Y = I$	k or e
3	GRSDA	$Tr(P^T X L_w X^T P)^{-1} Tr(P^T X L_b X^T P)$	None	k
4	L1/2-GLPCA	$\min_{U,V} \ X - UV^T\ _{1/2}^{1/2} + \beta tr(V^T L V)$	$V^T V = I$	k
5	p -GLPCA	$\min_{U,V} \ X - UV^T\ _p + \beta tr(V^T L V)$	$V^T V = I$	k
6	Proposed	$\min_{U,V} \ X - UV^T\ _{2,p}^2 + \beta tr(V^T L V) + \lambda_1 \ X - XW\ _F^2 + \lambda_2 \ W\ _1$	$V^T V = I$	No

is linearly separable from the other two classes, which are not linearly separable from each other. The Seeds dataset covers three different varieties of wheat—Kama, Rosa, and Canadian—each containing 70 elements, which were randomly selected for the experiment. The Soybean dataset has 19 classes, only the first 15 of which have been used in prior work. It contains 35 categorical attributes, some nominal and some ordered.

By using equation (5), $L_{2,p}$ norm can be applied in the SLPCA algorithm to improve the robustness. The previous research has established that $L_{2,p}$ -norm ($0 < p < 2$) performs better than other constraints [32]. To apply it, some parameters need to be set in advance, including the rescaling coefficient. Different p values are tried while other parameters are unchanged to find the optimal p value. The other parameters are then decided in the similar way. Figure 1 shows the distribution of 2D data after DR with the Iris dataset. In this experiment, $\alpha = 0.6$, $\lambda_1 = 0.5$, $\lambda_2 = 0.5$, $\rho = 1.0$, $p = 0.5$, and $\theta = 0.1$. For the original LPP, the neighborhood size, k , also needs to be determined, and $k = n_i = 1$ where n_i is the number of samples in the i -th class. It can be seen from the figure that the proposed SLPCA scheme represents the Iris data pattern more compactly than the other algorithms.

The clustering accuracies for the training set and test set of the schemes are compared in Table 2. The best results are highlighted in bold. Note that outside of a few exceptional cases, the proposed SLPCA scheme consistently displays the highest accuracy for both the training and test data, across the different datasets.

4.3. Experiments Using Image Datasets. In this section, the performances of the different schemes are compared with regard to image classification and reconstruction. The experiments were conducted on three face datasets: CMU PIE, Extended Yale-B, and ORL. For each image, salt and pepper noise were added, with a noise density of 0.01.

4.3.1. CMU PIE Dataset. The CMU PIE dataset is publicly available at <http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html>. It has 41,368 images of 68 individuals, with 4 expressions, 13 poses, and 43 illumination conditions for each person, and all the images have different illuminations and expressions. In our experiment, the Pose C09 subset was chosen, which contains 1632 images of 68 people, with 24 different images of each person. In the experiment, all the images are in gray scale and normalized to a resolution of 64×64 pixels for the sake of efficient computation. 70% of the CMU PIE images are used as the training dataset. The rest forms the test dataset.

4.3.2. ORL Face Dataset. The ORL face dataset consists of 400 images of 40 individuals, with 10 different images per person. It is publicly available at <http://www.face-rec.org/databases/>. The images of each person were taken at different times with different facial expressions and details and under varying lighting conditions. In the experiment, all the images are in gray scale and are manually cropped to 64×64 pixels. For this dataset, 60% of the images per subject were randomly chosen to form a training set. The remaining images

FIGURE 1: The distribution of 2D data after DR. (a) PCA. (b) LPP. (c) GRSDA. (d) L1/2-GLPCA. (e) p -GLPCA. (f) Proposed.

are then used as the test set. For each algorithm, the number of iterations is set to 20 and the best recognition rate is selected along with the corresponding residual error.

4.3.3. Yale-B Face Dataset. The Yale-B face dataset is also popular in face recognition research. It is publicly available

at <http://vision.ucsd.edu/~leekc/ExtYaleDatabase/Yale%20Face%20Database.htm>. This dataset consists of 2432 face images of 38 subjects under various lighting conditions. 64 gray scale images were selected for each subject, and each image was resized to 32×32 pixels. In this experiment, 10, 20, and 30 images per subject were chosen to form the

TABLE 2: The comparison of clustering accuracy for the training data (trr) and test data (tsr).

Data	%	PCA	LPP	GRSDA	L1/2-GLPCA	p -GLPCA	SLPCA
Iris	trr	58.16	59.76	58.84	60.47	58.98	61.33
	tsr	48.08	52.83	54.55	53.52	52.22	56.67
Soybean	trr	75.69	67.86	71.43	75.49	75.00	76.13
	tsr	71.24	68.42	68.42	70.85	70.44	71.25
Seeds	trr	54.47	56.09	54.74	57.00	56.27	58.42
	tsr	50.68	59.46	50.68	57.16	56.36	56.73

FIGURE 2: The comparison of the reconstructed images with different schemes.

TABLE 3: The comparison of average accuracy for the image datasets.

Data	%	L1/2-GLPCA	p -GLPCA		SLPCA	
			$p = 1$	$p = 0.5$	$p = 1$	$p = 0.5$
EYale-B	trr	65.24	68.81	74.88	76.17	77.44
	tsr	68.58	70.98	72.23	68.49	69.99
ORL face	trr	57.57	58.09	63.25	66.22	66.67
	tsr	64.00	69.46	64.95	69.53	70.80
CMU_PIE	trr	56.13	65.36	67.86	77.00	77.74
	tsr	67.25	64.18	68.43	70.62	70.70

training set, with the remaining images being formed as the test dataset. The same procedure was applied as with the ORL dataset.

To illustrate the data representation aspect of SLPCA, ORL face dataset is applied here. In this experiment, $\alpha = 1.0$, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\rho = 0.7$, and $\theta = 0.9$. Figure 2 shows the original images (in 1st column) and reconstructed images of L1/2-GLPCA, p -GLPCA, and SLPCA. The 2nd column shows the reconstruction images of L1/2-GLPCA while the 3rd and 4th column are the results of p -GLPCA at $p = [1, 0.5]$. The last three columns contain the reconstructed images of SLPCA with $p = [1, 0.5]$. Only the images of three ORL persons (out of 40 persons) are shown due to

space limitation. Observe from the figure that the images of a person with the proposed SLPCA scheme tend to be more similar than those with the other methods. This indicates that the class structures are generally more apparent with the SLPCA representation, and this motivates us to use it for data clustering and classification.

Table 3 compares the clustering accuracy for the GLPCA and SLPCA schemes for different p values. Here, L1/2-GLPCA and p -GLPCA used an $L_{2,1}$ -norm and a p -norm for the PCA aspects, respectively. SLPCA used an $L_{2,p}$ -norm. It can be seen from the table that SLPCA (with $p = 0.5$) achieved the best performance out of all the schemes. In this experiment, $\alpha = 1.0$, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\rho = 0.7$, $p = 0.5$, and

FIGURE 3: The comparison of the reconstruction and residual error rate for the different schemes.

FIGURE 4: The comparison of the error rates of the different schemes with varying θ values.

FIGURE 5: The comparison of the error rates with different image datasets.

$\theta=0.9$. Figure 3 compares the schemes with regard to the reconstruction error and residual error ($\|X - UVT\|F/\|X\|F$) when the subspace dimension was changed for the 30% CMU PIE training dataset. It can be seen that SLPCA with a proper p value produced the smallest reconstruction and residual error rate.

Figure 4 compares the error rates for different θ values. This reveals that SLPCA has a lower reconstruction error for any value of p and the lowest residual error when $p = 1.0$, except for p -GLPCA where $p = 0.5$. Note that in the case of SLPCA, the reconstruction error decreases as θ becomes larger, while the residual error increases slowly. Figure 5 shows a comparison of the schemes across the three different image datasets. It can be seen that 20% CMU PIE and ORL datasets have the lowest and largest reconstruction error rate, respectively. It seems that there is no relationship between the two types of error rates for the given datasets. Here, $\alpha = 1.0$, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\rho = 0.5$, $p = 0.5$, and $\theta = 0.5$.

From the experiment results given above, SLPCA turns out to be the most robust for handling DR for various types of data because it retains both the local and global structure of the data.

5. Conclusion

In this paper, a robust global-local scheme called SLPCA has been proposed for dimensionality reduction. It increases robustness against noise and more effectively preserves the global and local structure of the samples. In seeking a trade-off between complexity and efficiency, robust $L_{2,p}$ -norm-based PCA (R2P-PCA) was introduced for the GDR elements, while joint sparse representation and LPP (SR-LPP) was used for the LDR. In addition, the SR-LPP algorithm is parameter-free so as to avoid the difficulty of determining the neighborhood size. This allows the graph construction and weight assignment to be finished in one step. SR-LPP can also learn a projection matrix and a sparse similarity matrix simultaneously and adaptively. Experimental results show that the proposed scheme is capable of more accurate classification and better representation than other typical DR schemes.

In future work, the proposed scheme will be extended to take nonlinear settings into account. As a linear model cannot capture nonlinear distortion, a nonlinear model is required to obtain effective and reliable results. The identification of nonlinear models requires more data and involves more comprehensive analysis than their linear counterparts. At present, the performance of the representation has been studied only in linear settings. How to make the linear models work in nonlinear settings is another issue that requires further investigation. Analytical models deciding the values of the parameters employed in a DR scheme also need to be developed.

Data Availability

All data included in this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partly supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2016-0-00133, Research on Edge Computing via Collective Intelligence of Hyperconnection IoT Nodes), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute for Information and Communications Technology Promotion) (No. 2015-0-00914) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2017R1A2B2009095, Research on SDN-based WSN Supporting Real-time Stream Data Processing and Multiconnectivity, and 2019R1I1A1A01058780, Efficient Management of SDN-based Wireless Sensor Network Using Machine Learning Technique), the second Brain Korea 21 PLUS project.

References

- [1] J. H. Friedman, "On bias, variance, 0/1-loss, and the curse-of-dimensionality," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 55–77, 1997.
- [2] J. Thioulouse, D. Chessel, and S. P. Champely, "Multivariate analysis of spatial patterns: a unified approach to local and global structures," *Environmental and Ecological Statistics*, vol. 2, no. 1, pp. 1–14, 1995.
- [3] A. Anagnostopoulos, M. Vlachos, M. Hadjieleftheriou, E. Keogh, and P. S. Yu, "Global distance-based segmentation of trajectories," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Philadelphia, PA, USA, pp. 34–43, August 2006.
- [4] S. Günnemann, I. Färber, K. Virochsiri, and T. Seidl, "Subspace correlation clustering: finding locally correlated dimensions in subspace projections of the data," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 352–360, ACM, Beijing, China, August 2012.
- [5] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 5, pp. 952–965, 2015.
- [6] X. Lu and Y. Yuan, "Hybrid structure for robust dimensionality reduction," *Neurocomputing*, vol. 124, pp. 131–138, 2014.
- [7] X. Liu, "Global and local structure preservation for feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 6, pp. 1083–1095, 2013.
- [8] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?" in *Proceedings of the 2011 International Conference on Computer Vision*, IEEE, Barcelona, Spain, November 2011.
- [9] G. Banjac, P. Goulart, B. Stellato, and S. Boyd, "Infeasibility detection in the alternating direction method of multipliers for convex optimization," *Journal of Optimization Theory and Applications*, vol. 183, no. 2, pp. 490–519, 2019.

- [10] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou, "L2, 1-norm regularized discriminative feature selection for unsupervised learning," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, Barcelona, Spain, June 2011.
- [11] Bo Jiang, C. Ding, and J. Tang, "Graph-Laplacian PCA: closed-form solution and robustness," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* Portland, OR, USA, June 2013.
- [12] C. Ding, D. Zhou, X. He, and H. Zha, "R 1-PCA: rotational invariant L 1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 281–288, ACM, Pittsburgh, PA, USA, June 2006.
- [13] X. Chang, F. Nie, Y. Yang, C. Zhang, and H. Huang, "Convex sparse PCA for unsupervised feature learning," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 1, p. 3, 2016.
- [14] M. Li, H. Xi, and X. Zhu, "An incremental version of L-MVU for the feature extraction of MI-EEG," *Computational Intelligence and Neuroscience*, p. 2019, 2019.
- [15] D. K. Saxena, J. A. Duro, A. Tiwari, K. Deb, and Q. Zhang, "Objective reduction in many-objective optimization: linear and nonlinear algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 77–99, 2012.
- [16] C. Wei, J. Chen, and Z. Song, "Multilevel MVU models with localized construction for monitoring processes with large scale data," *Journal of Process Control*, vol. 67, pp. 176–196, 2018.
- [17] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016.
- [18] W. Chen, S. Gou, X. Wang, X. Li, and L. Jiao, "Classification of PolSAR images using multilayer autoencoders and a self-paced learning approach," *Remote Sensing*, vol. 10, no. 1, p. 110, 2018.
- [19] X. Ma, H. Wang, and J. Geng, "Spectral-spatial classification of hyperspectral image based on deep auto-encoder," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4073–4085, 2016.
- [20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [21] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [22] X. He, "Neighborhood preserving embedding," in *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, vol. 1, IEEE, Beijing, China, October 2005.
- [23] W. Yu, R. Wang, F. Nie, F. Wang, Q. Yu, and X. Yang, "An improved locality preserving projection with ℓ_1 -norm minimization for dimensionality reduction," *Neurocomputing*, vol. 316, pp. 322–331, 2018.
- [24] Y. Shao, "Supervised global-locality preserving projection for plant leaf recognition," *Computers and Electronics in Agriculture*, vol. 158, pp. 102–108, 2019.
- [25] Y. Lu, "Low-rank 2-D neighborhood preserving projection for enhanced robust image representation," *IEEE Transactions on Cybernetics*, vol. 49, no. 5, pp. 1859–1872, 2018.
- [26] Y. Lu, "Low-rank preserving projections," *IEEE Transactions on Cybernetics*, vol. 46, no. 8, pp. 1900–1913, 2015.
- [27] Y. Lu, "Robust flexible preserving embedding," *IEEE Transactions on Cybernetics*, 2019.
- [28] Y. Lu, C. Yuan, W. Zhu, and X. Li, "Structurally incoherent low-rank nonnegative matrix factorization for image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5248–5260, 2018.
- [29] P. P. Markopoulos, M. Dhanaraj, and A. Savakis, "Adaptive L1-norm principal-component analysis with online outlier rejection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1131–1143, 2018.
- [30] T. Zhang, Z. Tang, and X. Shen, "A fast generalized low rank representation framework based on $\ell_{2,p}$ norm minimization for subspace clustering," *IEEE Access*, vol. 5, pp. 23299–23311, 2017.
- [31] M. Zhao, M. Lin, B. Chiu, Z. Zhang, and X.-S. Tang, "Trace ratio criterion based discriminative feature selection via l2-norm regularization for supervised learning," *Neurocomputing*, vol. 321, pp. 1–16, 2018.
- [32] Q. Wang, Q. Gao, X. Gao, and F. Nie, " $\ell_{2,p}$ -Norm based PCA for image recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1336–1346, 2017.
- [33] S. Vishveshwara, K. V. Brinda, and N. Kannan, "Protein structure: insights from graph theory," *Journal of Theoretical and Computational Chemistry*, vol. 1, no. 1, pp. 187–211, 2002.
- [34] X. Zhu, J. Lafferty, and R. Rosenfeld, *Semi-supervised Learning with Graphs*, Dissertation, Carnegie Mellon University, Language Technologies Institute, School OF Computer Science, Pittsburgh, PA, USA, 2005.
- [35] Ke Huang and S. Aviyente, "Sparse representation for signal classification," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2007.
- [36] W. Zhang, P. Kang, X. Fang, L. Teng, and N. Han, "Joint sparse representation and locality preserving projection for feature extraction," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 7, pp. 1731–1745, 2019.
- [37] L. Zhuang, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, RI, USA, June 2012.
- [38] S. Lou, X. Zhao, Y. Chuang, H. Yu, and S. Zhang, "Graph regularized sparsity discriminant analysis for face recognition," *Neurocomputing*, vol. 173, pp. 290–297, 2016.
- [39] C.-M. Feng, "Joint-norm constraint and graph-Laplacian PCA method for feature extraction," *BioMed Research International*, vol. 201714 pages, 2017.
- [40] C.-M. Feng, Y.-L. Gao, J.-X. Liu, C.-H. Zheng, and J. Yu, "PCA based on graph Laplacian regularization and P-norm for gene selection and clustering," *IEEE Transactions on Nanobiotechnology*, vol. 16, no. 4, pp. 257–265, 2017.