A new embedding quality assessment method for manifold learning

Peng Zhang Member, IEEE, Yuanyuan Ren, and Bo Zhang

Abstract—Manifold learning is a hot research topic in the field of computer science. A crucial issue with current manifold learning methods is that they lack a natural quantitative measure to assess the quality of learned embeddings, which greatly limits their applications to real-world problems. In this paper, a new embedding quality assessment method for manifold learning, named as Normalization Independent Embedding Quality Assessment (NIEQA), is proposed. Compared with current assessment methods which are limited to isometric embeddings, the NIEQA method has a much larger application range due to two features. First, it is based on a new measure which can effectively evaluate how well local neighborhood geometry is preserved under normalization, hence it can be applied to both isometric and normalized embeddings. Second, it can provide both local and global evaluations to output an overall assessment. Therefore, NIEQA can serve as a natural tool in model selection and evaluation tasks for manifold learning. Experimental results on benchmark data sets validate the effectiveness of the proposed method.

Index Terms—Nonlinear Dimensionality reduction, Manifold learning, Data analysis

I. INTRODUCTION

A long with the advance of techniques to collect and store large sets of high-dimensional data, how to efficiently process such data issues a challenge for many fields in computer science, such as pattern recognition, visual understanding and data mining. The key problem is caused by “the curse of dimensionality” [1], that is, in handling with such data the computational complexities of algorithms often go up exponentially with the dimension.

The main approach to address this issue is to perform dimensionality reduction. Classical linear methods, such as Principal Component Analysis (PCA) [2], [3] and Multidimensional Scaling (MDS) [4], achieve their success under the assumption that data lie in a linear subspace. However, such assumption may not usually hold and a more realistic assumption is that data lie on or close to a low-dimensional manifold embedded in the high-dimensional ambient space. Recently, many methods have been proposed to efficiently find meaningful low-dimensional embeddings from manifold-modeled data, and they form a family of dimensionality reduction methods called manifold learning. Representative methods include Locally Linear Embedding (LLE) [5], [6], ISOMAP [7], [8], Laplacian Eigenmap (LE) [9], [10], Hessian LLE (HLLE) [11], Diffusion Maps (DM) [12], [13], Local Tangent Space Alignment (LTSA) [14], Maximum Variance Unfolding (MVU) [15], and Riemannian Manifold Learning (RML) [16].

Manifold learning methods have drawn great research interests due to their nonlinear nature, simple intuition, and computational simplicity. They also have many successful applications, such as motion detection [17], sample preprocessing [18], gait analysis [19], facial expression recognition [20], hyperspectral imagery processing [21], and visual tracking [22].

Despite the above success, a crucial issue with current manifold learning methods is that they lack a natural measure to assess the quality of learned embeddings. In supervised learning tasks such as classification, the classification rate can be directly obtained through label information and used as a natural tool to evaluate the performance of the classifier. However, manifold learning methods are fully unsupervised and the intrinsic degrees of freedom underlying high-dimensional data are unknown. Therefore, after training process, we can not directly assess the quality of the learned embedding. As a consequence, model selection and model evaluation are infeasible. Although visual inspection on the embedding may be an intuitive and qualitative assessment, it can not provide a quantitative evaluation. Moreover, it can not be used for embeddings whose dimensions are larger than three.

Recently, several approaches have been proposed to address the issue of embedding quality assessment for manifold learning, which can be cast into tow categories by their motivations.

- Methods based on evaluating how well the rank of neighbor samples, according to pairwise Euclidean distances, is preserved within each local neighborhood.
- Methods based on evaluating how well each local neigh-
These methods are proved to be useful to isometric manifold learning methods, such as ISOMAP and RML. However, a large variety of manifold learning methods output normalized embeddings, such as LLE, HILLE, LE, LTSA and MVU, just to name a few. In these methods, embeddings have unit variance up to a global scale factor. Then the distance rank of neighbor samples is disturbed in the embedding as pairwise Euclidean distances are no longer preserved. Meanwhile, anisotropic coordinate scaling caused by normalization can not be recovered by rigid motion. As a consequence, existent methods would report false quality assessments for normalized embeddings.

In this paper, we first propose a new measure, named Anisotropic Scaling Independent Measure (ASIM), which can efficiently compare the similarity between two configurations under rigid motion and anisotropic coordinate scaling. Then based on ASIM, we propose a novel embedding quality assessment method, named Normalization Independent Embedding Quality Assessment (NIEQA), which can efficiently assess the quality of normalized embeddings quantitatively. The NIEQA method owns three characteristics.

1) NIEQA can be applied to both isometric and normalized embeddings. Since NIEQA uses ASIM to assess the similarity between patches in high-dimensional input space and their corresponding low-dimensional embeddings, the distortion caused by normalization can be eliminated. Then even if the aspect ratio of a learned embedding is scaled, NIEQA can still give faithful evaluation of how well the geometric structure of data manifold is preserved.

2) NIEQA can provide both local and global assessments. NIEQA consists of two components for embedding quality assessment, a global one and a local one. The global assessment evaluates how well the skeleton of a data manifold, represented by a set of landmark points, is preserved, while the local assessment evaluates how well local neighborhoods are preserved. Therefore, NIEQA can provide an overall evaluation.

3) NIEQA can serve as a natural tool for model selection and evaluation tasks. Using NIEQA to provide quantitative evaluations on learned embeddings, we can select optimal parameters for a specific method and compare the performance among different methods.

In order to evaluate the performance of NIEQA, we conduct a series of experiments on benchmark data sets, including both synthetic and real-world data. Experimental results on these data sets validate the effectiveness of the proposed method.

The rest of the paper is organized as follows. A literature review on related works is presented in Section II. The Anisotropic Scaling Independent Measure (ASIM) is described in Section III. Then the Normalization Independent Embedding Quality Assessment (NIEQA) method is depicted in Section IV. Experimental results are reported in Section V. Some concluding remarks as well as outlooks for future research are given in Section VI.

II. LITERATURE REVIEW ON RELATED WORKS

In this section, the current state-of-the-art on embedding quality assessment methods are reviewed. For convenience and clarity of presentation, main notations used in this paper are summarized in Table I. Throughout the whole paper, all data samples are in the form of column vectors. The superscript of a data vector is the index of its component.

| Table I | MAIN NOTATIONS. |
|---------|-----------------|
| \( \mathbb{R}^n \) | \( n \)-dimensional Euclidean space where high-dimensional data samples lie |
| \( \mathbb{R}^m \) | \( m \)-dimensional Euclidean space, \( m < n \), where low-dimensional embeddings lie |
| \( x_i \) | The \( i \)-th data sample in \( \mathbb{R}^n \), \( i = 1, 2, \ldots, N \) |
| \( \mathcal{X} \) | \( \mathcal{X} = \{ x_1, x_2, \ldots, x_N \} \) |
| \( X \) | \( X = [ x_1 \ x_2 \ \cdots \ x_N ]^T, n \times N \) data matrix |
| \( X_i \) | \( X_i = [ x_{i1} \ x_{i2} \ \cdots \ x_{im} ], n \times k \) data matrix |
| \( \mathcal{N}_k(x_i) \) | The index set of the \( k \) nearest neighbors of \( x_i \) in \( \mathcal{X} \)|
| \( \mathcal{Y} \) | \( \mathcal{Y} = \{ y_1, y_2, \ldots, y_N \} \) |
| \( Y \) | \( Y = [ y_1 \ y_2 \ \cdots \ y_N ]^T, m \times N \) data matrix |
| \( Y_i \) | \( Y_i = [ y_{i1} \ y_{i2} \ \cdots \ y_{ik} ], m \times k \) data matrix |
| \( \mathcal{N}_k(y_i) \) | The index set of the \( k \) nearest neighbors of \( y_i \) in \( \mathcal{Y} \) |
| \( e_k \) | \( e_k = [ 1 \ 1 \ \cdots \ 1 ]^T, k \) dimensional column vector |
| \( \mathbb{I}_k \) | Identity matrix of size \( k \) |
| \( \| \cdot \|_2 \) | \( L_2 \) norm for a vector |
| \( \| \cdot \|_F \) | Frobenius norm for a matrix |

Goldberg and Ritov \[23\] proposed the Procrustes Measure (PM) that enables quantitative comparison of outputs of isometric manifold learning methods. For each \( \mathcal{X}_i \) and \( \mathcal{Y}_i \), their method first uses Procrustes analysis \[24\]–\[26\] to find an
optimal rigid motion transformation, consisting of a rotation and a translation, after which $Y_i$ best matches $X_i$. Then the local similarity is computed as

$$L(X_i, Y_i) = \sum_{j=1}^{k} \| x_{ij} - R y_{ij} - b \|_2^2,$$

where $R$ and $t$ are the optimal rotation matrix and translation vector, respectively. Finally, the assessment is is given by

$$M_P = \frac{1}{N} \sum_{i=1}^{N} L(X_i, Y_i) / \| H_k X_i \|_F^2,$$  \hspace{1cm} (1)

where $H_k = I_k - e_k e_k^T$.

An $M_P$ close to zero suggests a faithful embedding. Reported experimental results show that the PM method provides good estimation of embedding quality for isometric methods such as ISOMAP. However, as pointed out by the authors, PM is not suitable for normalized embedding since the geometric structure of every local neighborhood is distorted by normalization. Although a modified version of PM is proposed in [23], which eliminates global scaling of each neighborhood, it still can not address the issue of separate scaling of coordinates in the low-dimensional embedding.

Besides the PM method, a series of works follow the line that a faithful embedding would yield a high degree of overlap between the neighbor sets of a data sample and of its corresponding embedding. Several works are proposed by using different ways to define the overlap degree. A representative one is the LC meta-criteria (LCMC) proposed by Chen and Buja [27], [28], which can serve as a diagnostic tool for measuring local adequacy of learned embedding. The LCMC assessment is defined as the sum of local overlap degree and given by

$$M_{LC} = \frac{1}{kN} \sum_{i=1}^{N} | \mathcal{N}_k(x_i) \cap \mathcal{N}_k(y_i) |.$$  \hspace{1cm} (2)

Venna and Kaski [29] proposed an assessment method which consists of two measures, one for trustworthiness and one for continuity, based on the change of indices of neighbor samples in $\mathbb{R}^n$ and $\mathbb{R}^m$ according to pairwise Euclidean distances, respectively. Aguirre et al. proposed an alternative approach for quantifying the embedding quality, by evaluating the possible overlaps in the low-dimensional embedding. Their assessment is used for automatic choice of the number of nearest neighbors for LLE [30] and also exploited in [31] to evaluate the embedding quality of LLE with optimal regularization parameter. Akkucuk and Carroll [32] independently developed the Agreement Rate (AR) metric which shares the same form to $M_{LC}$. Based on AR, they suggested another useful assessment method called corrected agreement rate, by randomly reorganize the indices of data in $\mathcal{Y}$. Also with AR, France and Carroll [33] proposed a method using the RAND index to evaluate dimensionality reduction methods.

Lee and Verleysen [34], [35] proposed a general framework, named co-ranking matrix, for rank-based criteria. The aforementioned methods, which are based on distance ranking of local neighborhoods, can all be cast into this unified framework. The block structure of the co-ranking matrix also provides an intuitive way to visualize the differences between distinct methods. In [36], they further extended their work to circumvent the global scale dependency.

The above assessments based on overlap degrees of neighborhoods are implemented in the same way: an embedding with good quality corresponds to a high value of the assessment. They work well for isometric embeddings since pairwise distances within each neighborhood are preserved. However, when the embedding is normalized, the neighborhood structure is distorted since pairwise distances are no longer kept. The overlap degree would be much lower than expected even if the embedding is of high quality under visual inspection.

B. Global approaches

Tenenbaum et al. [7] suggested to use the residual variance as a diagnostic measure to evaluate the embedding quality. Given $\mathcal{X}$ and $\mathcal{Y}$, the residual variance is computed by

$$M_{RV} = 1 - \rho^2(G_X, D_Y),$$  \hspace{1cm} (3)

where $\rho(G_X, D_Y)$ is the standard linear correlation coefficients taken over all entries of $G_X$ and $D_Y$. Here $G_X(i, j)$ is the approximated geodesic distance between $x_i$ and $x_j$ [7] and $D_Y(i, j) = \| y_i - y_j \|_2$. A low value of $M_{RV}$ close to zero indicates a good equality of the embedding.

The $M_{RV}$ measure was applied to choose the dimension of learned embedding for ISOMAP [7] and the optimal parameter for LLE [37]. Nevertheless, for a normalized embedding the geodesic distances are no longer preserved and the reliability of $M_{RV}$ may decrease in such case.

Dollár et al. [38] proposed a supervised method for model evaluation problem of manifold learning. They assume that there is a very large ground truth data set containing the training data. Pairwise geodesic distances are approximated within this set using ISOMAP, and the assessment is defined as the average error between pairwise Euclidean distances in the embedding and corresponding geodesic distances. However, in real situations we do not usually have such ground truth set and their assessment can not be used in general cases.

Recently, Meng et al. proposed a new quality assessment criterion to encode both local-neighborhood-preserving and
global-structure-holding performances for manifold learning. In their method, a shortest path tree is first constructed from the \( k \)-NN neighborhood graph of training data. Then the global assessment is computed by using Spearman’s rank order correlation coefficient defined on the rankings of branch lengths. Finally, the overall assessment is defined to be a linear combination of the global assessment and \( M_{LC} \). In their work, normalization is treated as a negative aspect in quality assessment, while our work is to define a new assessment which is independent of normalization.

### III. ASIM: Anisotropic Scaling Independent Measure

In this section, we introduce a novel measure, named Anisotropic Scaling Independent Measure (ASIM), which can effectively evaluate the similarity between two configurations under rigid motion and anisotropic coordinate scaling. A synthetic example is first given in Subsection III-A to demonstrate why existent assessments fail under normalization. Then the motivation and overall description of ASIM are presented in Subsection III-B. Finally, the computational details are stated in Subsection III-C.

#### A. A synthetic example

We randomly generate 100 points within the area \([-2, 2] \times [-1, 1]\) in \( \mathbb{R}^2 \), which form the input data set \( X = \{x_1, x_2, \ldots, x_{100}\} \). Next we normalize \( X \) to get output data \( Y \) such that \( YY^T = I_2 \), which are taken as the embedding of \( X \). In fact, \( X \) can be obtained from \( Y \) through a rotation and anisotropic coordinate scaling, that is, \( X = RSY \) where

\[
R = \begin{pmatrix}
-0.9991 & 0.0434 \\
0.0434 & 0.9991
\end{pmatrix}, \quad S = \begin{pmatrix}
11.6414 & 0 \\
0 & 5.6236
\end{pmatrix}.
\]

In Fig. 1(a), \( x_i, i = 1, 2, \ldots, 100 \) are marked with blue dots and the 10 nearest neighbors of the origin in \( X \) are marked with blue circles. In Fig. 1(b), \( y_i, i = 1, 2, \ldots, 100 \) are marked with red dots and the 10 nearest neighbors of the origin in \( Y \) are marked with red squares. Meanwhile, the corresponding embeddings of the 10 nearest neighbors of the origin in \( X \) are marked with blue circles. From Fig. 1(b) we can see that the neighborhood of the origin change a lot after normalization. Only 6 nearest neighbors are still in the neighborhood after normalization and the overlap degree is only 60%. Meanwhile, we also compute the Procrustes measure \( M_P \) between \( X \) and \( Y \) and show it in Fig. 1(b). After normalization, \( M_P \) is as high as 0.8054.

Through this synthetic example, we can clearly observe the distortion on \( M_P \) and local neighborhood overlap degree caused by normalization.

#### B. Motivation and description of ASIM

Since a manifold is a topological space which is locally equivalent to a Euclidean subspace, an embedding would be faithful if it preserves the structure of local neighborhoods. Then we face a question that how to define the “preservation” of local neighborhood structure.

Under the assumption that the data manifold is dense, each local neighborhood \( X_i \) can be roughly viewed as a linear subspace embedded in the ambient space. Considering possible normalization on \( Y \), a rational and reasonable choice is to define a new measure which can efficiently assess the similarity between \( X_i \) and \( Y_i \) under rigid motion and anisotropic coordinate scaling.

Formally, for each index \( i \), we assume that there exists a rigid motion and anisotropic coordinate scaling between \( X_i \) and \( Y_i \). Since a rigid motion can be decomposed into a rotation...
and a translation, then for any $x_{ij} \in X_i$ we assume that
\[ x_{ij} = P_i D_i y_{ij} + t_i , \tag{4} \]
where $P_i \in \mathbb{R}^{n \times m}$ is orthogonal, that is, $P_i^T P_i = I_m$. $D_i$ is a diagonal matrix of rank $m$ and $t_i \in \mathbb{R}^n$ stands for an arbitrary translation.

To evaluate how similar $X_i$ and $Y_i$ are, our goal is to find optimal $P_i^*, D_i^*$ and $t_i^*$ such that $Y_i$ best matches $X_i$ under Eq. (4). Equivalently, we need to solve the following constrained optimization problem
\[
\begin{align*}
\min & \sum_{j=1}^k \| x_{ij} - P_i^* D_i^* y_{ij} - t_i^* \|_2^2 \\
\text{s.t.} & \quad P_i^T P_i = I_m \\
& \quad D_i \in \mathcal{D}(m),
\end{align*}
\tag{5}
\]
where $\mathcal{D}(m)$ is the set of all diagonal matrices of rank $m$.

Then the neighborhood “preservation” degree can be defined as the sum of squared distances between corresponding samples in $X_i$ and $Y_i$ under the above transformation. Formally, the anisotropic scaling independent measure (ASIM) is defined as follows
\[
M_{asim}(X_i, Y_i) = \sum_{j=1}^k \| x_{ij} - P_i^* D_i^* y_{ij} - t_i^* \|_2^2 / \sum_{j=1}^k \| x_{ij} \|_2^2 ,
\tag{6}
\]
or in matrix form
\[
M_{asim}(X_i, Y_i) = \| X_i - P_i^* D_i^* Y_i - t_i^* e_k^T \|_F^2 / \| X_i \|_F^2 ,
\tag{7}
\]
where the normalization item in denominator is introduced to eliminate arbitrary scaling.

C. Computation of ASIM

The optimization problem Eq. (5) does not admit a closed-form solution. Alternatively, we use gradient descent method to solve Eq. (5). Note that all $n \times m$ orthogonal matrices form the so-called Stiefel manifold, which is a Riemannian submanifold embedded in $\mathbb{R}^{nm}$. We denote this manifold by $St(n, m)$. Also note that $\mathcal{D}(m)$ is closed for matrix addition, multiplication and scalar multiplication, hence $\mathcal{D}(m)$ is homeomorphic to $\mathbb{R}^m$. Then Eq. (5) can be resolved by using gradient descent method over matrix manifolds.

For convenience of presentation, we first introduce the $\delta$ operator [39], which is defined as follows

**Definition 1.** When the $\delta$ operator is defined on a $n \times n$ square matrix $A = (a_{ij})$, $\delta(A)$ is a $n$-dimensional vector formed by the diagonal entries of $v$, that is,
\[ \delta(A) = (a_{11}, a_{22}, \ldots, a_{nn})^T . \]

When the $\delta$ operator is defined on a $n \times n$ square matrix $A = (a_{ij})$, $\delta(A)$ is a $n$-dimensional vector formed by the diagonal entries of $v$, that is,
\[ \delta(A) = (a_{11}, a_{22}, \ldots, a_{nn})^T . \]

The $\delta$ operator can be compounded, which yields
\[
\delta^2(v) = v
\]
\[
\delta^2(A) = \begin{pmatrix} a_{11} \\ a_{22} \\ \vdots \\ a_{nn} \end{pmatrix}.
\]

With the above notations, Eq. (5) now can be rewritten in matrix form as
\[
\begin{align*}
\min_{P_i, D_i, t_i} & \quad \| X_i - P_i D_i Y_i - t_i e_k^T \|_F^2 \\
\text{s.t.} & \quad P_i \in St(n, m), \quad D_i \in \mathcal{D}(m) .
\end{align*}
\tag{8}
\]

Next, we solve Eq. (8) in three steps, which are described respectively as follows.

1) Computation $t_i^*$: Let $L_i = P_i D_i$ and note that for any matrix $A$, $\| A \|_F^2 = \text{tr} (A^T A)$. Then the objective function can be written as
\[
f(L_i, t_i) = \text{tr} \left( (X_i - L_i Y_i - t_i e_k^T)^T (X_i - L_i Y_i - t_i e_k^T) \right) .
\tag{9}
\]

By using the propositions of matrix trace Eq. (9) can be expanded as
\[
f(L_i, t_i) = \text{tr} \left( X_i^T X_i \right) + \text{tr} \left( Y_i^T L_i^T L_i Y_i \right) - 2 \text{tr} \left( Y_i^T X_i \right) + 2 \text{tr} \left( e_k^T X_i t_i \right) + \text{tr} \left( e_k^T Y_i^T L_i^T t_i \right) .
\tag{10}
\]

Taking derivative with respect to $t_i$ yields
\[
\frac{\partial f(L_i, t_i)}{\partial t_i} = 2kt_i - 2X_i e_k + 2L_i Y_i .
\]

Since $f(L_i, t_i)$ is a strict convex function of $t_i$, then by making both sides of the above equation to be zero, we can get the optimal solution to $t_i$ as follows
\[ t_i^* = \frac{1}{k} (X_i - P_i D_i Y_i) e_k . \tag{11} \]

Substitute $t_i^*$ into Eq. (8), and the latter one is rewritten as
\[
\begin{align*}
\min_{P_i, D_i} & \quad \| \bar{X}_i - P_i D_i \bar{Y}_i \|_F^2 \\
\text{s.t.} & \quad P_i \in St(n, m), \quad D_i \in \mathcal{D}(m) ,
\end{align*}
\tag{12}
\]
where $\bar{X}_i = X_i (I_k - \frac{1}{k} e_k e_k^T)$ and $\bar{Y}_i = Y_i (I_k - \frac{1}{k} e_k e_k^T)$. 
2) Computation of \(D^*_i\): In the second step, we compute the optimal solution \(D^*_i\) to \(D_i\) with respect to \(P_i\). Let \(A_i = \bar{Y}_i\bar{Y}_iT\) and \(B_i = P_iX_i\bar{Y}_iT\), and denote the objective function in Eq. (12) by \(f(P_i, D_i)\). Then we have
\[
f(P_i, D_i) = \text{tr} \left( D^2A_i \right) - 2 \text{tr} \left( D(B_i) \right) + \text{tr} \left( \bar{X}_i\bar{X}_iT \right) \\
= \sum_{j=1}^m a^{(i)}_{jj} b^{(i)}_{jj} \frac{1}{a^{(i)}_{jj}} - 2 \sum_{j=1}^m b^{(i)}_{jj} \frac{1}{a^{(i)}_{jj}} + \text{tr} \left( \bar{X}_i\bar{X}_iT \right),
\]
where \(a^{(i)}_{jj}\), \(b^{(i)}_{jj}\), and \(d^{(i)}_{jj}\) are the \(j\)-th diagonal entries of matrices \(A_i\), \(B_i\), and \(D_i\), respectively.

Since \(a^{(i)}_{jj} \geq 0\), \(j = 1, 2, \ldots, m\), \(f\) is a convex function of vector \(\delta(D_i)\). Taking partial derivative with respect to \(d^{(i)}_{jj}\) \((j = 1, 2, \ldots, m)\) and by making them to be zero, we can get the global optimal solutions to \(d^{(i)}_{jj}\) \((j = 1, 2, \ldots, m)\) as follows
\[
d^{(i)}_{jj} = \frac{b^{(i)}_{jj}}{a^{(i)}_{jj}}, \quad j = 1, 2, \ldots, m.
\]

Then \(D^*_i\) is given by
\[
D^*_i = (\delta^2(A_i))^{-1}\delta^2(B_i).
\]

Substituting Eq. (13) into \(f\) yields
\[
f(P_i) = \text{tr} \left( (\delta^2(A_i))^{-1} \delta^2(B_i) \right) - 2 \text{tr} \left( (\delta^2(A_i))^{-1} \delta^2(B_i) \right) + \text{tr} \left( \bar{X}_i\bar{X}_iT \right) \\
= \sum_{j=1}^m \frac{a^{(i)}_{jj} b^{(i)}_{jj}}{(a^{(i)}_{jj})^2} - 2 \sum_{j=1}^m \frac{b^{(i)}_{jj} a^{(i)}_{jj}}{(a^{(i)}_{jj})^2} + \text{tr} \left( \bar{X}_i\bar{X}_iT \right) \\
= - \sum_{j=1}^m \frac{b^{(i)}_{jj}}{a^{(i)}_{jj}} + \text{tr} \left( \bar{X}_i\bar{X}_iT \right).
\]

Let \(M_i = \bar{X}_i\bar{Y}_iT (\delta^2(A_i))^{-1/2}\), then \(f(P_i)\) can be rewritten as
\[
f(P_i) = - \sum_{j=1}^m (P_i^T M_i)^2 + \text{tr} \left( \bar{X}_i\bar{X}_iT \right) \\
= - \text{tr} \left( (P_i^T M_i) \odot (P_i^T M_i) \right) + \text{tr} \left( \bar{X}_i\bar{X}_iT \right),
\]
where \(P_{ij}\) and \(M_{ij}\) are the \(j\)-th columns of matrices \(P_i\) and \(M_i\), respectively. \(\odot\) stands for the Hadamard product over matrices. The optimization problem Eq. (12) can be transformed into
\[
\max_{P_i} \phi(P_i) = \text{tr} \left( (P_i^T M_i) \odot (P_i^T M_i) \right) \quad \text{s.t.} \quad P_i \in \text{St}(n,m).
\]

3) Computation of \(P^*_i\): In the third step, we use gradient descent method over matrix manifold to solve Eq. (14), which is an optimization problem for matrix function over the Stiefel manifold \(\text{St}(n,m)\).

Denote the gradient of \(\phi\) in \(\mathbb{R}^{nm}\) by \(\nabla \phi(P_i)\) and the gradient of \(\phi\) on \(\text{St}(n,m)\) by \(\nabla \phi(P_i)\), then by the proposition of Stiefel manifold [40]. \(\nabla \phi(P_i)\) is the projection of \(\nabla \delta \phi(P_i)\) onto the tangential space at \(P_i\) and can be computed by the following formula
\[
\nabla \phi(P_i) = \nabla \delta \phi(P_i) - P_i \frac{\partial^2 \delta \phi(P_i)}{\partial P_i^T} P_i + P_i \frac{\partial^2 \delta \phi(P_i)}{\partial P_i} P_i, \tag{15}
\]

Now all we need is to compute \(\nabla \delta \phi(P_i)\). Let \(F(P_i) = (P_i^T M_i) \odot (P_i^T M_i)\). From matrix calculus, the differentiation of \(\phi\) with respect to \(P_i\) is
\[
D\phi(P_i) = (\text{vec} M_i)^T DF(P_i), \tag{16}
\]
where the vec operator reformulates a \(n \times m\) matrix into a \(nm\)-dimensional vector by stacking its columns one underneath other.

Next we derive \(DF(P_i)\). First, we have
\[
dF(P_i) = 2(M_i^T P_i) \odot (M_i^T dP_i) = 2W_m ((M_i^T P_i) \odot (M_i^T dP_i)) W_m,
\]
where \(\odot\) stands for the Kronecker product over matrices and \(W_m = (\text{vec} w_1 w_1^T, \text{vec} w_2 w_2^T, \ldots, \text{vec} w_m w_m^T)\) is an \(m^2 \times m\) matrix. \(w_i, i = 1, 2, \ldots, m\) is an \(m\)-dimensional vector who has 1 in its \(i\)-th component and 0 elsewhere. Then we have
\[
\text{vec} dF(P_i) = 2 \text{vec}(W_m ((M_i^T P_i) \odot (M_i^T dP_i)) W_m) \\
= 2(W_m \odot W_m) \text{vec}(M_i^T P_i \odot (M_i^T dP_i)) \\
= 2 \text{vec}(W_m \odot W_m) (H_i \odot I_m) \text{vec}(M_i^T P_i),
\]
where \(H_i = ((I_m \odot K_{mm}) ((\text{vec} M_i^T P_i) \odot (I_m)) \odot I_m\). Here \(K_{mm}\) is a permutation matrix of order \(m^2\), and for any square matrix \(M\) of order \(m\), \(K_{mm} \text{vec} M = \text{vec} M^T\). Then by matrix calculus [41], we have
\[
DF(P_i) = 2(W_m \odot W_m) (H_i \odot I_m) (I_m \odot M_i^T).
\]

Furthermore, through algebraic deduction and Eq. (16), we have
\[
D\phi(P_i) = (\text{vec} M_i)^T DF(P_i) = 2 \text{vec}(M_i \delta^2(P_i^T M_i))^T.
\]

Then \(\nabla \phi(P_i)\) is given by the following formula
\[
\nabla \phi(P_i) = 2M_i \delta^2(P_i^T M_i),
\]
and by using Eq. (15), \(\nabla \phi(P_i)\) now reads
\[
\nabla \phi(P_i) = 2M_i \delta^2(P_i^T M_i) - P_i P_i^T \delta^2(P_i^T M_i) - P_i \delta^2(P_i^T M_i) M_i^T P_i. \tag{17}
\]

Given a step length for iteration, we apply gradient descent method to find \(P^*_i\) such that \(\nabla \phi(P_i)\) vanishes. In each iteration, we first update \(P_i\) as
\[
\tilde{P}_i = P_i + \alpha \nabla \phi(P_i).
\]

Then we retract \(\tilde{P}_i\) to \(\text{St}(n,m)\). From the property of \(\text{St}(n,m)\), such retraction can be obtained through the QR
decomposition of $\tilde{P}_i$. Let $\tilde{P}_i = Q_i R_i$, where $Q_i \in St(n, m)$ and $R_i$ is an upper-triangular matrix. The retraction of $\tilde{P}_i$ to $St(n, m)$ is just $Q_i$.

In each iteration, we use $Q_i$ to update $P_i$ until $\|\nabla \phi(P_i)\|_F$ is less than a given threshold $\epsilon$. After $P_i^*$ is computed, $D_i^*$ can be given by Eq. (13), and the optimal value to Eq. (12) is $f(P_i^*, D_i^*)$.

4) The algorithm and discussion: Finally, we summarize the computation process of $M_{asim}$ in Algorithm 1.

When the dimension $n$ of input samples is very high, performing QR decomposition of $\tilde{P}_i$ in each iteration will greatly increase of computational complexity of Algorithm 1. A possible solution to this issue is first projecting $X_i$ to its tangential space, denoted as $TX_i$, and then computing $M_{asim}(TX_i, Y_i)$. When data are densely distributed on the manifold, $TX_i$ can optimally recover the local linear structure of a manifold. Therefore, such strategy is feasible. The tangential space can be approximated by using PCA, MDS or the method proposed in [42].

IV. NORMALIZATION INDEPENDENT EMBEDDING QUALITY ASSESSMENT

When assessing the quality of embeddings, we need to consider both local and global evaluations. This leads to two issues.

- Does the embedding preserve the global topology of the manifold?
- Does the embedding preserve the geometric structure of local neighbor neighborhoods?

In this section, we propose Normalization Independent Embedding Quality Assessment method (NIEQA) to address these two issues, which is independent of normalization. NIEQA is based on the ASIM measure stated in Section III and consists of two assessments, a local one and a global one. In the following subsections, we introduce these two assessments respectively as well as how NIEQA can be implemented in model selection and model evaluation.

A. Local assessment

For local neighborhood $X_i$ on a data manifold and its corresponding low-dimensional embedding $Y_i$, the local measure $M_{asim}$ defined in last section characterizes how well local neighborhood structure is preserved and is independent of normalization. Therefore, we define the local assessment as the mean value of $M_{asim}(X_i, Y_i)$ over index $i$, that is,

$$M_L(X, Y) = \frac{1}{N} \sum_{i=1}^{N} M_{asim}(X_i, Y_i).$$

B. Global assessment

From geometric intuition, if an embedding preserves the global topology of the data manifold well, then such embedding should preserve relative positions among “representative” samples on the manifold. In other words, if we treat these “representative” samples as a local neighborhood, where pairwise Euclidean distances among neighborhood samples are replaced with pairwise geodesic distances on the manifold, then a good embedding should preserve the geometric structure of this neighborhood.

Motivated by the above consideration, we define the global assessment as the matching degree between the aforementioned described neighborhood and its corresponding embedding under rigid motion and anisotropic coordinate scaling. The computation of the global assessment consists of three steps, which are depicted below, respectively.

1) Selecting landmark points. First, for each training sample $x_i$, find its $k_l$ nearest neighbors. Treat $x_i$ as a node in a graph and add edges among neighboring samples with edge length being pairwise Euclidean distance. Through such construction we get a connected graph. Then we use the shortest path length between $x_i$ and $x_j$ to approximate the geodesic distance between them for all $i$ and $j$. Next, we count how many shortest paths going through each $x_i$ and record this number as its importance degree. Finally, the top 10% most
important data samples are selected as landmark points on the manifold and the set they formed is denoted by \(X_l\).

2) **Computing** \(\hat{Y}_l\). Once \(X_l\) is fixed in the first step, the distance between any two landmark points is defined to be the approximated geodesic distance. Then we implement MDS [4] to \(X_l\) to obtain its isometric embedding \(\hat{Y}_l\), which optimally preserve relative positions of landmark points on the manifold. Note that the dimensions of \(\hat{Y}_l\) and \(Y_l\) are equal, and the latter one is the subset in \(Y\) corresponding to \(X_l\).

3) **Computing the global assessment.** We define the global assessment \(M_G\) to be the ASIM measure between \(\hat{Y}_l\) and \(Y_l\)

\[
M_G(X, Y) = M_{asim}(\hat{Y}_l, Y_l), \tag{19}
\]

where \(\hat{Y}_l\) and \(Y_l\) are the \(m \times l\) data matrices corresponding to \(\hat{Y}_l\) and \(Y_l\), respectively.

**Remark 1.** During landmark points selection, the parameter \(k_l\) needs to be set manually. Based on experimental experience, setting \(k_l = 0.1N\) can yield a connected graph that approximates the manifold structure well. However, if the graph is disconnected under current \(k_l\), \(k_l\) should be set to be the smallest integer which makes the graph fully connected.

The landmark points selection method stated above has intuitive geometric motivation and is easy to implement. It can also be replaced with other more accurate yet more complicated approaches, for example, the methods proposed in [23] and [42].

C. Implementation in model evaluation and model selection

In this subsection, we state how to implement the NIEQA method to model evaluation and model selection for manifold learning.

- **Model evaluation.** Given \(X\), suppose that we have two embeddings, namely \(Y_1\) and \(Y_2\), obtained by different manifold learning methods. Then we say that \(Y_1\) owns better locality preservation than \(Y_2\) if \(M_L(X, Y_1) < M_L(X, Y_2)\) and vice versa. We say \(Y_1\) owns better global topology preservation than \(Y_2\) if \(M_G(X, Y_1) < M_G(X, Y_2)\) and vice versa.

- **Model selection.** Given \(X\) and a set of parameters \(\mathcal{P} = \{p_1, p_2, \ldots, p_l\}\), for each parameter \(p_i\) we compute its corresponding embedding \(Y^{(i)}\) using specific manifold learning method. Then we use \(M_G\) or \(M_L\) or their combination, which depends on the user’s demand, to evaluate the quality of \(Y^{(i)}\). Finally, the \(p_i\) corresponding to the lowest assessment score is chosen to be the optimal parameter.

V. Experiments

In this section, the effectiveness of the NIEQA method is validated through a series of experimental tests on benchmark data sets. In Subsection V-A, NIEQA is applied to model evaluation for manifold learning. In Subsection V-A, NIEQA is used to select optimal parameters for the LLE method which outputs normalized embeddings. In experiments, NIEQA is compared with three commonly used assessment methods. We compute \(1 - M_{LC}\) instead of \(M_{LC}\) to obtain a unified criterion, that is, a small assessment value close to zero indicates good quality of the embedding. The benchmark data sets used in experiments are briefly depicted in Table II and notations for methods are summarized in Table III.

A. Model evaluation

In the first experiment, we apply NIEQA to model evaluation of the Swissroll manifold with parameter equation

\[
\begin{align*}
x_1 &= u^1 \
&= u^1 \cos u^1 \\
x_2 &= u^2 \\
x_3 &= u^1 \sin u^1
\end{align*}
\]

We use LLE [5], LE [10], LTSA [14], ISOMAP [7] and RML [16] to learn this manifold, respectively. 1000 training samples are randomly generated and the number of nearest

---

### Table II

**Description of experimental data sets.**

| Data manifold    | \(N\) | \(n\) | \(m\) | Description                                      |
|------------------|------|------|------|-------------------------------------------------|
| Swissroll        | 1000 | 3    | 2    | Surface isometrically embedded in \(\mathbb{R}^3\) |
| Swishole         | 1000 | 3    | 2    | Surface embedded in \(\mathbb{R}^3\)            |
| Gaussian         | 1000 | 3    | 2    | Surface isometrically embedded in \(\mathbb{R}^3\) |
| Ileface          | 1493 | 560  | 2    | Face manifold with resolution 28 \times 20      |

### Table III

**Notations used in experiments.**

| Notation | Description                                                                 |
|----------|-----------------------------------------------------------------------------|
| \(M_P\)  | Procrustes measure (Eq. [1]) \[23\]                                         |
| \(M_P^{\ast}\) | \(M_P\) with global scaling removed \[23\]                                   |
| \(M_{LC}\) | LMC measure (Eq. [2]) \[27\], \[28\]                                       |
| \(M_{RV}\) | Residual Variance measure (Eq. [1]) \[7\]                                   |
| \(M_L\)  | Local assessment of NIEQA (Eq. [18])                                       |
| \(M_G\)  | Global assessment of NIEQA (Eq. [19])                                       |
| \(M_t\)  | Matching degree between \(Y\) and ground truth \(U\), \(M_{asim}(Y, U)\)   |
neighbors is 10. Figs. 2(c)-(g) shows the results of manifold learning, where \( \mathcal{X} \) and \( \mathcal{U} \) stands for the training data and the groundtruth of intrinsic degrees of freedom, respectively. By visual inspection, the embeddings given by LTSA and RML are the most similar to \( \mathcal{U} \). The one given by ISOMAP is a litter worse, and the one learned by LLE has a great change in global shape. LE fails to recover the geometric structure of \( \mathcal{U} \).

For embeddings given by the above methods, we compute the different assessments described in Table III and use bar plots to visualize their values in Figs. 2(h)-(m). From the bar plots, we can see that \( M_P \) only works for isometric embeddings given by ISOMAP and RML while reports false high values for normalized embeddings learned by LTSA and LLE. Although \( M_P^\nu \) eliminates the affects of global scaling, only the scale of \( M_P \) is normalized and it still reports false high values for normalized embeddings. \( M_{LC} \) and \( M_{RV} \) fails to output reasonable equality evaluations. It should be noted that \( M_{RV} \) is originally designed for the ISOMAP method and hence works well for the embedding given by ISOMAP.

The two assessments \( M_L \) and \( M_G \) in NIEQA provide overall and reasonable evaluations on embedding quality for various methods. \( M_L \) shows that LTSA and RML best preserve local neighborhood. LLE and ISOMAP perform worse, and LE performs the worst. \( M_G \) further indicates that the global-shape-preservation of the embedding given by LLE is not good. This completely matches visual inspection, which demonstrates that NIEQA can effectively evaluate the quality of both isometric and normalized embeddings.

Besides, the bar plot of the matching degree \( M_t \) between \( \mathcal{Y} \) and \( \mathcal{U} \) is shown in Fig. 2(n). We can see that only \( M_L \) and \( M_G \) match \( M_t \), which validates the effectiveness of NIEQA.

Similar to the first experiment, we apply NIEQA to model evaluation of the Swisshole manifold, which shares the same parameter equation to Swisshole. The difference is that the set of intrinsic degree of freedoms \( \mathcal{U} \) is no longer a convex set, where a rectangular region in \( \mathcal{U} \) is digged out. Therefore, Swisshole manifold is geodesic non-connected. 1000 training samples are randomly generated from the manifold and the number of nearest neighbors \( k \) is 10. The learned low-dimensional embeddings and the bar plots of quality assessments are shown in Fig. 3.

From Fig. 3 we can see that LTSA and RML correctly learned the geometric structure of \( \mathcal{U} \) with the highest quality over other approaches. The embedding given by LLE has a distortion in global shape. ISOMAP and LE fails to learn the structure of \( \mathcal{U} \). From the bar plots in Figs. 3(h)-(l), we can see that \( M_L \) reports a reasonable quality assessment and matches \( M_t \) well which is illustrated in Fig. 5(m). \( M_P \) and \( M_P^\nu \) works only for isometric embeddings provided by ISOMAP and RML. \( M_{LC} \) and \( M_{RV} \) fails to report reasonable evaluations. Since Swisshole manifold is geodesic non-connected, using shortest path length would fail to approximate geodesic distance. Therefore, we do not compute the global assessment \( M_G \) in NIEQA.

In the third experiment, we apply NIEQA to model evaluation of the Gaussian manifold, whose parameter equation is

\[
\begin{align*}
    x^1 &= u^1 \\
    x^2 &= u^2 \\
    x^3 &= \frac{1}{2\pi} \exp\left\{-\left((u^1)^2 + (u^2)^2\right)/2\right\}
\end{align*}
\]

1000 training samples are randomly generated from the manifold and the number of nearest neighbors \( k \) is 10. Fig. 4 shows the learned low-dimensional embeddings as well as bar plots of different quality assessments.

From Fig. 4 we can observe that except LE all the other methods successfully learned the geometric structure of this manifold, whilst the quality of the embedding given by ISOMAP is a litter worse. From Figs. 3(h)-(m), we can see that \( M_P^\nu \) performs well in this case by eliminating the global scaling factor. This is due to the isotropic property of this manifold. \( M_{LC} \) reports correct evaluations but still leans against to RML. \( M_{RV} \) fails to assess the embeddings correctly. Both the two assessments in NIEQA successfully evaluate the quality of different embeddings and match \( M_t \) well. Note that the Gaussian surface is isotropic, hence the measure \( M_P^\nu \) also works. However, for anisotropic surfaces like Swissroll and Swisshole, only removing global scaling would not yield a reasonable assessment.

In the next experiment, we apply NIEQA to model evaluation tasks on the lifeface data set, which is a high-dimensional image manifold. As the code of RML on high-dimensional data is not available, we do not test RML on this data set. The training data contain 1965 face images, and the intrinsic degrees of freedom are the angle of face orientation and the variation of facial emotion. We randomly select 1493 images as training data such that the data graph constructed via ISOMAP is connected. We apply LLE, LE, ISOMAP and LTSA to learn this manifold with 15 nearest neighbors. The two dimensional embeddings learned by these methods and bar plots of the quality assessments given by different methods are shown in Fig. 5.

From Fig. 5 we can see that the embedding given by LLE does not recover the change of face orientation. The other methods all successfully extract the two intrinsic degrees of freedom despite the difference in embedding shape. The
above visual inspection is also validated by the bar plots of quantitative assessments shown in Figs. 5(e)-(i). $M_{LC}$, $M_{RV}$ and $M_L$ all suggest that the quality of the embedding given by LLE is poor, while the others are almost of the same quality. $M_{LC}$ and $M_L$ indicate that the embedding given by LTSA is of the highest quality. $M_P$ and $M_P^r$ fail in this case.

**Remark 2.** In experiments on high-dimensional image manifold, we did not compute $M_G$. The reason lies in that the computation of $M_G$ needs to estimate geodesic distances based on shortest graph paths. However, we have no prior knowledge on the underlying geometric structure of image manifolds, hence using $M_G$ to assess the global topology would yield unknown bias. Also note that the values of intrinsic degrees of freedom for image manifolds are unknown, hence we do not compute $M_I$ either.

### B. Model selection

In this subsection, we take the LTSA method as an example to demonstrate the application of NIEQA to model selection task. The most important parameter for LTSA is the number of nearest neighbors $k$. We first apply NIEQA to selecting $k$...
for LTSA on the Swissroll data set. Similar to the first experiment in Section V-A. We randomly select 1000 samples from the Swissroll manifold as training data. The values of $k$ are chosen to be integers from 5 to 24. For each $k$, an embedding is learned with LTSA, which are shown in Fig. 6. The assessments given by NIEQA corresponding to different values of $k$ are shown in Fig. 8(a). From the figure we can see that when $k$ is taking values between 6 and 15, LTSA would produce embeddings with high quality. This observation is also supported by visual inspection from Fig. 6, which validates the effectiveness of the NIEQA method.

In the second experiment, we apply NIEQA to select optimal $k$ for LTSA on the lleface data set. Training data are the same to those used in the experiment in Section V-A. Values of $k$ are taken to be integers from 5 to 24. For each $k$, an embedding is learned with LTSA, which is shown in Fig. 7. Corresponding quality assessment given by NIEQA are illustrated in Fig. 8(b), from which we can see that the embedding corresponding to $k = 14$ is of the highest quality. We can also observe that when $k > 8$, the quality of embeddings improves along with the increase of $k$, which is also validated by visual inspections from Fig. 7.
VI. CONCLUSIONS AND DISCUSSIONS

In this paper, we proposed a novel normalization independent embedding quality assessment (NIEQA) method for manifold learning, which has wider application range than current approaches. We first propose a new local measure, which can quantitatively evaluate how well local neighborhood structure is preserved under rigid motion and anisotropic coordinate scaling. Then the NIEQA method, which is designed based on this new measure, can effectively and quantitatively evaluate the quality of both isometric and normalized embeddings. Furthermore, the NIEQA method considers both local and global topology, thus it can yield an overall assessment. Experimental tests on benchmark data sets validate the effectiveness of the proposed method.

Some discussions and possible improvements in future works are stated below.

- The measure $M_{asim}$ is computed by using gradient descent method on matrix manifold. Whether the solution converges to a global optima remains unproved and is the key part of our future works. Meanwhile, we will also consider how to design more efficient iteration method.
Fig. 5. Manifold learning results on **lleface**. (a)-(d) Embeddings learned by various methods. The name of each method is stated below each subfigure. (e)-(i) Bar plots of different assessments on learned embeddings. The lower-case character under each bar corresponds to the index of the subfigure above.
Fig. 6. Embeddings given by LTSA on Swissroll data set with different values of $k$.

Fig. 7. Embeddings given by LTSA on face data set with different values of $k$. 
to accelerate convergence.

- The NIEQA method is based on a local matching methodology. Its basic assumption is that the manifold is densely sampled and training data strictly lie on the manifold. For data manifold with noise or outliers, the efficiency of NIEQA may be affected. A possible solution to this issue is to implement denoising or outlier removal process before training.

- Based on NIEQA, whether we can design a manifold learning method with better learning performance is also one of our future works.

ACKNOWLEDGEMENTS

This work was partly supported by the NNSF of China grant no. 90820007, and the 973 Program of China grant no. 2007CB311002.

REFERENCES

[1] D. L. Donoho, High-dimensional data analysis: the curses and blessings of dimensionality, in: Proceedings of American Mathematical Society Conference on Math Challenges of the 21st Century, 2000.

[2] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.

[3] I. T. Jolliffe, Principal Component Analysis, 2nd Edition, Springer, 2002.

[4] T. F. Cox, M. A. A. Cox, Multidimensional Scaling, Second Edition, Chapman & Hall, 2000.

[5] S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[6] L. K. Saul, S. T. Roweis, Think globally, fit locally: Unsupervised learning of low dimensional manifolds, Journal of Machine Learning Research 4 (2003) 119–155.

[7] J. B. Tenenbaum, V. Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.

[8] V. De Silva, J. B. Tenenbaum, Global versus local methods in nonlinear dimensionality reduction, in: Advances in Neural Information Processing Systems 15, Vol. 15, 2003, pp. 705–712.

[9] M. Belkin, Problems of learning on manifolds, Ph.D. thesis, The University of Chicago (2003).

[10] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373–1396.

[11] D. L. Donoho, C. Grimes, Hessian eigenmaps: Locally linear embedding, Science 290 (5500) (2000) 2319–2323.

[12] K. Weinberger, L. Saul, Unsupervised learning of image manifolds by semidefinite programming, International Journal of Computer Vision 70 (1) (2006) 77–90.

[13] T. Lin, H. Zha, Riemannian manifold learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (5) (2008) 796–809.

[14] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimensionality reduction via tangent space alignment, SIAM Journal on Scientific Computing 26 (1) (2005) 313–338.
Fig. 8. Graph plot of embedding quality assessments in model selection experiment for the LTSA method. (a) Assessments on Swissroll data set with different value of $k$. (b) Assessment $ML$ on face data set with different value of $k$.

Learning Research: Workshop and Conference proceedings, 2008, pp. 21–35.

[36] J. A. Lee, M. Verleysen, Scale-independent quality criteria for dimensionality reduction, Pattern Recognition Letters 31 (2010) 2248–2257.

[37] O. Kouropteva, O. Okun, M. Pietikäinen, Selection of the optimal parameter value for the locally linear embedding algorithm, in: The 1st International Conference on Fuzzy Systems and Knowledge Discovery, 2002, pp. 359–363.

[38] P. Dollár, V. Rabaud, S. Belongie, Non-isometric manifold learning: analysis and an algorithm, in: ICML ‘07: Proceedings of the 24th international conference on Machine learning, ACM, New York, NY, USA, 2007, pp. 241–248.

[39] J. Dattorro, Convex Optimization & Euclidean Distance Geometry, Meboo Publishing USA, 2005.

[40] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, NJ, USA, 2007.

[41] J. R. Magnus, H. Neudecker, Matrix differential calculus with applications in statistics and econometrics, 2nd Edition, John Wiley & Sons, 1999.

[42] P. Zhang, H. Qiao, B. Zhang, An improved local tangent space alignment method for manifold learning, Pattern Recognition Letters 32 (2) (2011) 181–190.

[43] V. d. Silva, J. B. Tenenbaum, Selecting landmark points for sparse manifold learning, in: Advances in Neural Information Processing Systems (NIPS), Vol. 18, 2006, pp. 1241–1248.

[44] J. Li, P. Hao, Finding representative landmarks of data on manifolds, Pattern Recogn. 42 (11) (2009) 2335–2352.