Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Pattern Recognition 39 (2006) 1799 – 1804
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Rapid and brief communication

The LLE and a linear mapping

F.C. Wu*, Z.Y. Hu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, People's Republic of China

Received 23 September 2005; accepted 30 March 2006

Abstract

The locally linear embedding (LLE) is considered an effective algorithm for dimensionality reduction. In this short note, some of its key properties are studied. In particular, we show that: (1) there always exists a linear mapping from the high-dimensional space to the low-dimensional space such that all the constraint conditions in the LLE can be satisfied. The implication of the existence of such a linear mapping is that the LLE cannot guarantee a one-to-one mapping from the high-dimensional space to the low-dimensional space for a given data set; (2) if the LLE is required to globally preserve distance, it must be a PCA mapping; (3) for a given high-dimensional data set, there always exists a local distance-preserving LLE. The above results can bring some new insights into a better understanding of the LLE.

Keywords: Locally linear embedding (LLE); Linear mapping; Principal component analysis (PCA)

1. Introduction

The locally linear embedding (LLE) is considered one of effective algorithms for dimensionality reduction [1]. It has been used to solve various problems in pattern recognition [2–5]. However, to our knowledge, the LLE has the following two problems to solve:

- If two data points $\{\mathbf{z}_i, \mathbf{z}_j\}$ in the high-dimensional space are different, their corresponding data points $\{\mathbf{y}_i, \mathbf{y}_j\}$ in a lower-dimensional space must be different.
- If $\{\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{ik}\}$ are the k-neighborhood of $\mathbf{z}_i$, then $\{\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{ik}\}$ must be the k-neighborhood of $\mathbf{y}_i$.

Since the LLE does not involve any metric, in addition, taking into account our following discussions, we think the above two problems cannot completely be solved without additional constraints being further imposed.

In this note, we will show that

- There always exists a linear mapping from the **z**-space to the **y**-space such that all the constraint conditions in the LLE can be satisfied.
- If the LLE is required to (globally) preserve distance, it must be a principal component analysis (PCA) mapping.
- For any given high-dimensional data set, there always exists a local distance-preserving LLE.

In the note, we suppose the reader is familiar with the algorithms such as the LLE, the PCA, etc. In addition, we suppose the reader is familiar with fundamentals of matrix analysis. Besides, in this note, neither simulations nor experiments are reported, the correctness of results lie in our proofs.

2. A linear mapping from the z-space to the y-space

The following proposition shows that there always exists a linear mapping from the high-dimensional **z**-space to the lower-dimensional **y**-space such that all the constraint conditions in the LLE can be satisfied.

* Corresponding author. Tel.: +86 10 826 14 521; fax: +86 10 625 51 993.
E-mail addresses: fcwu@nlpr.ia.ac.cn (F.C. Wu), huzy@nlpr.ia.ac.cn (Z.Y. Hu).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.03.019

Proposition 1. *Let* $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\} \subset R^n$ *be a high-dimensional data set*, $\mathbf{Z}_{n\times m} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m]$, *if weight matrix* $\mathbf{W}_{m\times m}$ *satisfies*

$$\mathbf{1}_m^{\mathrm{T}} \mathbf{W}_{m\times m} = \mathbf{0}_m^{\mathrm{T}}, \tag{1}$$

$$\mathbf{Z}_{n\times m} \mathbf{W}_{m\times m} = \mathbf{0}_{n\times m}, [1] \tag{2}$$

then $\forall d \leqslant r (= rank(\hat{\mathbf{Z}}_{n\times m}))$, *there exists always a linear mapping* $\mathbf{A}_{d\times n}$ *and a lower-dimensional data set* $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\} \subset R^d$ *such that* $\mathbf{Y}_{d\times m} = \mathbf{A}_{d\times n} \hat{\mathbf{Z}}_{n\times m}$, *and* $\mathbf{Y}_{d\times m}$ *satisfies all the constraint conditions in the LLE*:

$$\mathbf{Y}_{d\times m} \mathbf{W}_{m\times m} = \mathbf{0}_{d\times m}, \tag{3}$$

$$\mathbf{Y}_{d\times m} \mathbf{1}_m = \mathbf{0}_d, \tag{4}$$

$$\mathbf{Y}_{d\times m} \mathbf{Y}_{d\times m}^{\mathrm{T}} = \mathbf{I}_{d\times d}, \tag{5}$$

where

$$\mathbf{1}_m = (1, 1, \ldots, 1)^{\mathrm{T}}, \quad \mathbf{Y}_{d\times m} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m],$$

$$\hat{\mathbf{Z}}_{n\times m} = [\mathbf{z}_1 - \mathbf{z}_0, \mathbf{z}_2 - \mathbf{z}_0, \ldots, \mathbf{z}_m - \mathbf{z}_0], \quad \mathbf{z}_0 = \frac{1}{m} \sum_{j=1}^{m} \mathbf{z}_j.$$

Proof. Let $\mathbf{C}_{m\times m} = \hat{\mathbf{Z}}_{n\times m}^{\mathrm{T}} \hat{\mathbf{Z}}_{n\times m}$. Since $\hat{\mathbf{Z}}_{n\times m} \mathbf{1}_m = \mathbf{0}_n$, the 1-vector $\mathbf{1}_m$ is the eigenvector of the matrix $\mathbf{C}_{m\times m}$ for eigenvalue 0. Because $rank(\hat{\mathbf{Z}}_{n\times m}) = r$, $\hat{\mathbf{Z}}_{n\times m}$ could be decomposed by the SVD decomposition as

$$\hat{\mathbf{Z}}_{n\times m} = \mathbf{U}_{n\times r} \mathbf{\Sigma}_{r\times r} \mathbf{V}_{r\times m}, \tag{6}$$

where $\mathbf{U}_{n\times r}$ is a column-orthogonal matrix, $\mathbf{V}_{r\times m}$ a row-orthogonal matrix, and $\mathbf{\Sigma}_{r\times r}$ a diagonal matrix with positive diagonal elements. Let $\mathbf{V}_{r\times m} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m]$. Since

$$\begin{aligned}\hat{\mathbf{Z}}_{n\times m} \mathbf{W}_{m\times m} &= [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m] \mathbf{W}_{m\times m} \\ &\quad - [\mathbf{z}_0, \mathbf{z}_0, \ldots, \mathbf{z}_0] \mathbf{W}_{m\times m} \\ &= \mathbf{0}_{n\times m},\end{aligned}$$

by Eq. (6) we have

$$\mathbf{V}_{r\times m} \mathbf{W}_{m\times m} = \mathbf{0}_{r\times m}. \tag{7}$$

Since $\mathbf{V}_{r\times m}$ is a row-orthogonal matrix, and in addition each of its row vectors is orthogonal to the null space of matrix $\mathbf{C}_{m\times m}$, we have

$$\mathbf{V}_{r\times m} \mathbf{V}_{r\times m}^{\mathrm{T}} = \mathbf{I}_{r\times r}, \tag{8}$$

$$\mathbf{V}_{r\times m} \mathbf{1}_m = \mathbf{0}_r. \tag{9}$$

Let $\hat{\mathbf{A}}_{r\times n} = (\mathbf{\Sigma}_{r\times r})^{-1} (\mathbf{U}_{n\times r})^{\mathrm{T}}$, then from Eq. (6), we have

$$\mathbf{V}_{r\times m} = \hat{\mathbf{A}}_{r\times n} \hat{\mathbf{Z}}_{n\times m}. \tag{10}$$

$\forall d \leqslant r$, let $\mathbf{P}_{d\times r}$ be a row-orthogonal matrix, which defines a linear mapping from r-dimensional space to d-dimensional space, for example, let $\mathbf{P}_{d\times r} = [\mathbf{I}_{d\times d}, \mathbf{0}_{d\times(r-d)}]$, then

$$\mathbf{Y}_{d\times m} = \mathbf{P}_{d\times r} \mathbf{V}_{r\times m} = (\mathbf{P}_{d\times r} \hat{\mathbf{A}}_{r\times n}) \hat{\mathbf{Z}}_{n\times m}$$

and $\mathbf{Y}_{d\times m}$ satisfies the conditions (3)–(5). □

Remarks. 1. Since our mapping is a linear one, from a given data set $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\} \subset R^n$, we can always obtain its corresponding set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\} \subset R^d$ $(d < r)$, but a one-to-one correspondence cannot be guaranteed between the two sets. In other words, by the linear mapping, we can only guarantee that in the r-dimensional space, if two data points $\{\mathbf{z}_i, \mathbf{z}_j\}$ are different, their corresponding $\{\mathbf{y}_i, \mathbf{y}_j\} \subset R^r$ must be different, not in the $d(< r)$-dimensional space.

2. Since the constraint conditions on set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ obtained by this proposition are identical to those in the LLE, our linear mapping must be included in the LLE. By this reasoning, we think the two problems outlined at the beginning of this comment cannot be solved without further conditions being imposed.

3. The relationship between our linear mapping and the LLE

In Remark 2 of the above section, we indicate that our linear mapping must be included in the LLE. In this section, some specifics of the relationship between the linear mapping and the LLE will be provided.

Proposition 2. *Let* $N_l(\mathbf{W}_{m\times m})$ *be the left null space of* $\mathbf{W}_{m\times m}$. *If the dimension of* $N_l(\mathbf{W}_{m\times m})$, *denoted as* $\dim N_l(\mathbf{W}_{m\times m})$, *is of* $(r+1)$, *then the LLE must be our linear mapping*.

Proof. Since the row-orthogonal matrix $\begin{bmatrix} \mathbf{V}_{r\times m} \\ (1/\sqrt{m}) \mathbf{1}_m^{\mathrm{T}} \end{bmatrix}$ satisfies

$$\begin{bmatrix} \mathbf{V}_{r\times m} \\ (1/\sqrt{m})\, \mathbf{1}_m^{\mathrm{T}} \end{bmatrix} \mathbf{W}_{m\times m} = \mathbf{0}_{(r+1)\times m}, \tag{11}$$

by $\dim N_l(\mathbf{W}_{m\times m}) = r + 1$, each one of the row vectors of lower-dimensional data matrix $\mathbf{Y}_{d\times m}$ obtained by the LLE must be a linear combination of the row vectors of matrix $\mathbf{V}_{d\times m}$:

$$\mathbf{y}^i = \sum_{j=1}^{r} p_{ji} \mathbf{v}^j, \quad i = 1, 2, \ldots, d.$$

Hence,

$$\mathbf{Y}_{d\times r} = \mathbf{P}_{d\times r} \mathbf{V}_{r\times m}, \tag{12}$$

where $\mathbf{P}_{d\times r} = [p_{ij}]$, which defines a linear mapping from the r-dimensional space to d-dimensional space.

[1] In the LLE, the equality $\mathbf{Z}_{n\times m} \mathbf{W}_{m\times m} = \mathbf{0}_{n\times m}$ is implied. Otherwise, minimizing $\|\mathbf{Y}_{d\times m} \mathbf{W}_{m\times m}\|_F^2$ is meaningless. This is because if the equality does not hold, the obtained $\mathbf{Y}_{d\times m}$ from the minimization is not the best one in terms of local linearity preserving.

Since $\mathbf{Y}_{d\times m}\mathbf{Y}^{\mathrm{T}}_{d\times m}=\mathbf{I}_{d\times d}$, by Eq. (12) we have

$$\mathbf{P}_{d\times r}\mathbf{P}^{\mathrm{T}}_{d\times r}=\mathbf{P}_{d\times r}\mathbf{V}_{r\times m}\mathbf{V}^{\mathrm{T}}_{r\times m}\mathbf{P}^{\mathrm{T}}_{d\times r}=\mathbf{Y}_{d\times m}\mathbf{Y}^{\mathrm{T}}_{d\times m}=\mathbf{I}_{d\times d}.$$

Then, $\mathbf{P}_{d\times r}$ must be a row-orthogonal matrix. By Proposition 1, the LLE must be our linear mapping. □

If $\dim N_l(\mathbf{W}_{m\times m})=k+1$ $(k>r)$, then the lower-dimensional data matrix $\mathbf{Y}_{d\times m}$ obtained by the LLE will depend not only on the row vectors of matrix $\mathbf{V}_{d\times m}$ but other vectors also in the null space $N_l(\mathbf{W}_{m\times m})$, hence the LLE will not necessarily be a linear mapping. The following proposition gives the relationship between our linear mapping and the LLE in this case.

Proposition 3. *If* $\dim N_l(\mathbf{W}_{m\times m})=k+1$ $(k>r)$, *then the lower-dimensional data points* $\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_m\}$ *obtained by the LLE must satisfy*

$$\mathbf{Y}_{d\times m}=\mathbf{P}_{d\times k}\begin{bmatrix}\mathbf{V}_{r\times m}\\ \mathbf{Q}_{(k-r)\times m}\end{bmatrix}, \tag{13}$$

where $\mathbf{Q}_{(k-r)\times m}$ *is a row-orthogonal matrix satisfying*

$$\hat{\mathbf{Z}}_{n\times m}\mathbf{Q}^{\mathrm{T}}_{(k-r)\times m}=\mathbf{0}_{n\times(k-r)}. \tag{14}$$

Proof. Since

$$\begin{bmatrix}\mathbf{V}_{r\times m}\\ (1/\sqrt{m})\,\mathbf{1}^{\mathrm{T}}_m\end{bmatrix}\mathbf{W}_{m\times m}=\mathbf{0}_{(r+1)\times m}$$

and

$$\dim N_l(\mathbf{W}_{m\times m})=k+1,$$

we can always find out a row-orthogonal matrix $\mathbf{Q}_{(k-r)\times m}$ such that

$$\mathbf{V}_{r\times m}\mathbf{Q}^{\mathrm{T}}_{(k-r)\times m}=\mathbf{0}_{r\times(k-r)}$$

and

$$\begin{bmatrix}\mathbf{V}_{r\times m}\\ \mathbf{Q}_{(k-r)\times m}\\ (1/\sqrt{m})\,\mathbf{1}^{\mathrm{T}}_m\end{bmatrix}\mathbf{W}_{m\times m}=\mathbf{0}_{(k+1)\times m}.$$

Hence, the lower-dimensional data points $\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_m\}\subset R^d$ obtained by the LLE must satisfy

$$\mathbf{Y}_{d\times m}=\mathbf{P}_{d\times k}\begin{bmatrix}\mathbf{V}_{r\times m}\\ \mathbf{Q}_{(k-r)\times m}\end{bmatrix},$$

where $\mathbf{P}_{d\times r}$ is a row-orthogonal matrix. Since $\mathbf{V}_{r\times m}\mathbf{Q}^{\mathrm{T}}_{(k-r)\times m}=\mathbf{0}_{r\times(k-r)}$, by Eq. (6) we have

$$\hat{\mathbf{Z}}_{n\times m}\mathbf{Q}^{\mathrm{T}}_{(k-r)\times m}=\mathbf{0}_{n\times(k-r)}. \qquad \square$$

4. The distance-preserving LLE

In order to obtain a lower-dimensional data set $\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_m\}\subset R^d$ from a high-dimensional data set $\{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_m\}\subset R^n$, the constraint $\mathbf{Y}_{d\times m}\mathbf{Y}^{\mathrm{T}}_{d\times m}=\mathbf{I}_{d\times d}$, which is not related to the high-dimensional data, is a main constraint in the LLE. By Proposition 1, this constraint cannot guarantee the one-to-one correspondence and k-neighborhood relationship between the two data sets. In this section, we will use the constraint $\mathbf{Y}^{\mathrm{T}}_{d\times m}\mathbf{Y}_{d\times m}=\mathbf{C}_{m\times m}$ instead of $\mathbf{Y}_{d\times m}\mathbf{Y}^{\mathrm{T}}_{d\times m}=\mathbf{I}_{d\times d}$. We have the following proposition:

Proposition 4. *If the lower-dimensional data set* $\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_m\}\subset R^d$ *obtained by the LLE satisfies Eqs.* (3), (4) *and*

$$\mathbf{Y}^{\mathrm{T}}_{d\times m}\mathbf{Y}_{d\times m}=\mathbf{C}_{m\times m}, \tag{15}$$

then the LLE must be a PCA mapping.

Before giving a proof for this proposition, we need at first to prove the following lemma:

Lemma. *Let* $\dim N_l(\mathbf{W}_{m\times m})=k+1$ $(k\geqslant r)$ *and* $\mathbf{Q}_{k\times m}=\begin{bmatrix}\mathbf{V}_{r\times m}\\ \mathbf{Q}_{(k-r)\times m}\end{bmatrix}$[2] *then* $\mathbf{Y}_{d\times m}$ *is a lower-dimensional data matrix satisfying the constraints* (3) *and* (4) *if and only if there exists a linear mapping* $\mathbf{A}_{d\times k}$ *from k-dimensional space to d-dimensional space such that*

$$\mathbf{Y}_{d\times m}=\mathbf{A}_{d\times k}\mathbf{Q}_{k\times m}. \tag{16}$$

Proof. (⇒) Since the row-orthogonal matrix $\begin{bmatrix}\mathbf{Q}_{k\times m}\\ (1/\sqrt{m})\,\mathbf{1}^{\mathrm{T}}_m\end{bmatrix}$ satisfies

$$\begin{bmatrix}\mathbf{Q}_{k\times m}\\ (1/\sqrt{m})\mathbf{1}^{\mathrm{T}}_m\end{bmatrix}\mathbf{W}_{m\times m}=\mathbf{0}_{(k+1)\times m}$$

and $\dim N_l(\mathbf{W}_{m\times m})=k+1$, by the constraint (3), there must exist a matrix $\tilde{\mathbf{A}}_{d\times(k+1)}$ such that

$$\mathbf{Y}_{d\times m}=\tilde{\mathbf{A}}_{d\times(k+1)}\begin{bmatrix}\mathbf{Q}_{k\times m}\\ (1/\sqrt{m})\,\mathbf{1}^{\mathrm{T}}_m\end{bmatrix}. \tag{17}$$

Then, from the constraint (4), we have

$$\frac{1}{\sqrt{m}}\mathbf{Y}_{d\times m}\mathbf{1}_m=\tilde{\mathbf{A}}_{d\times(k+1)}\begin{bmatrix}\mathbf{Q}_{k\times m}\\ (1/\sqrt{m})\,\mathbf{1}^{\mathrm{T}}_m\end{bmatrix}\left(\frac{1}{\sqrt{m}}\mathbf{1}_m\right)$$
$$=\tilde{\mathbf{A}}_{d\times(k+1)}\begin{bmatrix}\mathbf{0}_k\\ 1\end{bmatrix}=\mathbf{0}_d.$$

Thus, $\tilde{\mathbf{A}}_{d\times(k+1)}$ must be of the following form:

$$\tilde{\mathbf{A}}_{d\times(k+1)}=[\mathbf{A}_{d\times k},\mathbf{0}_d].$$

[2] When $k=r$, the matrix $\mathbf{Q}_{(k-r)\times m}$ will disappear in this expression.

Therefore, substituting it into Eq. (17) gives

$$\mathbf{Y}_{d\times m} = \mathbf{A}_{d\times k}\mathbf{Q}_{k\times m}.$$

($\Leftarrow$) Since the row-orthogonal matrix $\begin{bmatrix} \mathbf{Q}_{k\times m} \\ (1/\sqrt{m})\,\mathbf{1}_m^{\mathrm{T}} \end{bmatrix}$ satisfies

$$\begin{bmatrix} \mathbf{Q}_{k\times m} \\ (1/\sqrt{m})\,\mathbf{1}_m^{\mathrm{T}} \end{bmatrix} \mathbf{W}_{m\times m} = \mathbf{0}_{(k+1)\times m},$$

by Eq. (16), the matrix $\mathbf{Y}_{d\times m}$ satisfies the constraints (3) and (4). □

By the lemma and the discussions in Section 2, the matrix $\mathbf{Y}_{d\times m}$ satisfies the constraints (3)–(5) if and only if the matrix $\mathbf{A}_{d\times k}$ in Eq. (16) is a row-orthogonal matrix.

Next, we will prove Proposition 4.

Proof of Proposition 4. Let $\dim N_l(\mathbf{W}_{m\times m}) = k+1$ ($k \geqslant r$), and $\mathbf{Y}_{d\times m}$ be a lower-dimensional data matrix satisfying the constraints (3) and (4), then by the lemma, we have

$$\mathbf{Y}_{d\times m} = \mathbf{A}_{d\times k}\mathbf{Q}_{k\times m}.$$

Therefore, by Eqs. (6) and (15), we have

$$\mathbf{Q}_{k\times m}^{\mathrm{T}}\mathbf{A}_{d\times k}^{\mathrm{T}}\mathbf{A}_{d\times k}\mathbf{Q}_{k\times m} = \mathbf{C}_{m\times m} = \hat{\mathbf{Z}}_{n\times m}^{\mathrm{T}}\hat{\mathbf{Z}}_{n\times m} = \mathbf{V}_{r\times m}^{\mathrm{T}}\mathbf{\Sigma}_{r\times r}^{2}\mathbf{V}_{r\times m}.$$

From the row orthogonality of $\mathbf{Q}_{k\times m}$, we have

$$\mathbf{A}_{d\times k}^{\mathrm{T}}\mathbf{A}_{d\times k} = \begin{bmatrix} \mathbf{\Sigma}_{r\times r}^{2} & \\ & \mathbf{0}_{(k-r)\times(k-r)} \end{bmatrix}. \tag{18}$$

Hence, $d = r = rank(\hat{\mathbf{Z}}_{n\times m})$,[3] and

$$\mathbf{A}_{d\times k} = \mathbf{R}_{d\times d}[\mathbf{\Sigma}_{d\times d}, \mathbf{0}_{r\times(k-d)}], \tag{19}$$

where $\mathbf{R}_{d\times d}$ is an orthogonal matrix.

By Eqs. (16), (19) and (6), we have

$$\hat{\mathbf{Z}}_{n\times m} = \mathbf{U}_{n\times d}\mathbf{R}_{d\times d}^{\mathrm{T}}\mathbf{Y}_{d\times m}. \tag{20}$$

Then, we can obtain

$$\mathbf{Y}_{d\times m} = \mathbf{R}_{d\times d}\mathbf{U}_{n\times d}^{\mathrm{T}}\hat{\mathbf{Z}}_{n\times m}. \tag{21}$$

This is just a PCA of $\hat{\mathbf{Z}}_{n\times m}$. □

According to Eq. (20), the LLE satisfying the constraint (15) must preserve distance, i.e. $\forall \mathbf{z}_i, \mathbf{z}_j$ and their corresponding $\mathbf{y}_i, \mathbf{y}_j$, the following equality holds:

$$\sqrt{(\mathbf{z}_i - \mathbf{z}_j)^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_j)} = \sqrt{(\mathbf{y}_i - \mathbf{y}_j)^{\mathrm{T}}(\mathbf{y}_i - \mathbf{y}_j)}.$$

[3] If $\mathbf{Y}_{d\times m}^{\mathrm{T}}\mathbf{Y}_{d\times m} = \mathbf{C}_{m\times m}$, then $d = rank(\hat{\mathbf{Z}}_{n\times m})$, i.e. the dimensionality of $\mathbf{z}$ can only be reduced to $rank(\hat{\mathbf{Z}}_{n\times m})$.

This is because

$$\begin{aligned}(\mathbf{z}_i - \mathbf{z}_j)^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_j) &= (\hat{\mathbf{z}}_i - \hat{\mathbf{z}}_j)^{\mathrm{T}}(\hat{\mathbf{z}}_i - \hat{\mathbf{z}}_j) \\ &= ((\mathbf{y}_i - \mathbf{y}_j)^{\mathrm{T}}\mathbf{R}_{d\times d}\mathbf{U}_{n\times d}^{\mathrm{T}}) \\ &\quad \times (\mathbf{U}_{n\times d}\mathbf{R}_{d\times d}^{\mathrm{T}}(\mathbf{y}_i - \mathbf{y}_j)) \\ &= (\mathbf{y}_i - \mathbf{y}_j)^{\mathrm{T}}(\mathbf{y}_i - \mathbf{y}_j).\end{aligned}$$

Hence, the constraint (15) ensures that there is a one-to-one correspondence and the k-neighborhood keeps unchanged between the two data sets. On the other hand, if the LLE preserves distance, then it must satisfy the constraint (15), and thus we have the following corollary.

Corollary. *The (global) distance-preserving LLE must be a PCA mapping.*

5. The local distance-preserving LLE

According to the discussions in Section 4, the (global) distance-preserving LLE must be a PCA mapping. In this section, we have an investigation on the local distance-preserving LLE.

Definition. Let $\{\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_s}\}$ be the s-neighborhood of $\mathbf{z}_i$, and

$$d_{i_l}^2 = (\mathbf{z}_{i_l} - \mathbf{z}_i)^{\mathrm{T}}(\mathbf{z}_{i_l} - \mathbf{z}_i), \quad l = 1, 2, \ldots, s, \quad i = 1, 2, \ldots, m. \tag{22}$$

If the lower-dimensional data set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\} \subset R^d$ obtained by the LLE satisfies the constraints (3) and (4), and additionally

$$(\mathbf{y}_{i_l} - \mathbf{y}_i)^{\mathrm{T}}(\mathbf{y}_{i_l} - \mathbf{y}_i) = d_{i_l}^2, \quad l = 1, 2, \ldots, s, \quad i = 1, 2, \ldots, m, \tag{23}$$

then the LLE is called the local distance-preserving LLE.

Proposition 5. *For any given high-dimensional data set* $\{\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_s}\}$, *we can always find out a matrix* $\mathbf{A}_{d\times k}$, *such that* $\mathbf{Y}_{d\times m} = \mathbf{A}_{d\times k}\mathbf{Q}_{k\times m}$ *is a lower-dimensional data matrix satisfying the local distance-preserving constraint.*

Proof. By the lemma in Section 3, if a matrix $\mathbf{Y}_{d\times m}$ satisfies the constraints (3) and (4), then we have

$$\mathbf{Y}_{d\times m} = \mathbf{A}_{d\times k}\mathbf{Q}_{k\times m}, \tag{24}$$

where the matrix $\mathbf{A}_{d\times k}$ is unknown. Therefore, we only need to determine the matrix $\mathbf{A}_{d\times k}$ by the constraint (23).

Let $\mathbf{Q}_{k\times m} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m]$, then, $\mathbf{y}_i = \mathbf{A}_{d\times k}\mathbf{q}_i$, $i = 1, 2, \ldots, m$. By (23), we have

$$\tilde{\mathbf{q}}_{i_l}^{\mathrm{T}}\mathbf{A}_{d\times k}^{\mathrm{T}}\mathbf{A}_{d\times k}\tilde{\mathbf{q}}_{i_l} = d_{i_l}^2, \quad l = 1, 2, \ldots, s, \quad i = 1, 2, \ldots, m, \tag{25}$$

where $\tilde{\mathbf{q}}_{i_l} = \mathbf{q}_{i_l} - \mathbf{q}_i = [q_1^{(i_l)}, q_2^{(i_l)}, \ldots, q_k^{(i_l)}]^{\mathrm{T}}$.

Let

$$\mathbf{X}_{k\times k} = \mathbf{A}_{d\times k}^{\mathrm{T}}\mathbf{A}_{d\times k} \equiv [x_{ij}],$$

$$\mathbf{x}_{k(k+1)/2} = [x_{11}, x_{12}, \ldots, x_{1k}, x_{22}, x_{23}, \ldots, x_{2k}, \ldots, x_{(k-1)(k-1)}, x_{(k-1)k}, x_{kk}]^{\mathrm{T}},$$

$$\mathbf{B}_{s\times[k(k+1)/2]}^{(i)} = \begin{bmatrix} (\tilde{q}_1^{(i_1)})^2 & 2\tilde{q}_1^{(i_1)}\tilde{q}_2^{(i_1)} & \cdots & 2\tilde{q}_1^{(i_1)}\tilde{q}_k^{(i_1)} & \cdots & (\tilde{q}_{k-1}^{(i_1)})^2 & 2\tilde{q}_{(k-1)}^{(i_1)}\tilde{q}_k^{(i_1)} & (\tilde{q}_k^{(i_1)})^2 \\ (\tilde{q}_1^{(i_2)})^2 & 2\tilde{q}_1^{(i_2)}\tilde{q}_2^{(i_2)} & \cdots & 2\tilde{q}_1^{(i_2)}\tilde{q}_k^{(i_2)} & \cdots & (\tilde{q}_{k-1}^{(i_2)})^2 & 2\tilde{q}_{(k-1)}^{(i_2)}\tilde{q}_k^{(i_2)} & (\tilde{q}_k^{(i_2)})^2 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ (\tilde{q}_1^{(i_s)})^2 & 2\tilde{q}_1^{(i_s)}\tilde{q}_2^{(i_s)} & \cdots & 2\tilde{q}_1^{(i_s)}\tilde{q}_k^{(i_s)} & \cdots & (\tilde{q}_{k-1}^{(i_s)})^2 & 2\tilde{q}_{(k-1)}^{(i_s)}\tilde{q}_k^{(i_s)} & (\tilde{q}_k^{(i_s)})^2 \end{bmatrix},$$

$$\mathbf{b}_s^{(i)} = [d_{i_1}^2, d_{i_2}^2, \ldots, d_{i_s}^2]^{\mathrm{T}}.$$

Then, the constraint (25) can be rewritten as

$$\mathbf{B}_{s\times[k(k+1)/2]}^{(i)}\mathbf{x}_{k(k+1)/2} = \mathbf{b}_s^{(i)}, \quad i = 1, 2, \ldots, m.$$

Hence, we have

$$\mathbf{B}_{ms\times[k(k+1)/2]}\mathbf{x}_{k(k+1)/2} = \mathbf{b}_{ms}, \tag{26}$$

where

$$\mathbf{B}_{ms\times[k(k+1)/2]} = \begin{bmatrix} \mathbf{B}_{s\times[k(k+1)/2]}^{(1)} \\ \mathbf{B}_{s\times[k(k+1)/2]}^{(2)} \\ \vdots \\ \mathbf{B}_{s\times[k(k+1)/2]}^{(m)} \end{bmatrix}, \quad \mathbf{b}_{ms} = \begin{bmatrix} \mathbf{b}_s^{(1)} \\ \mathbf{b}_s^{(2)} \\ \vdots \\ \mathbf{b}_s^{(m)} \end{bmatrix}.$$

Since the local distance-preserving constraint is weaker than the above-mentioned (global) distance-preserving constraint, Eq. (26) must have solution. Here we use the least squares solution:

$$\mathbf{x}_{k(k+1)/2} = \mathbf{B}_{ms\times[k(k+1)/2]}^{+}\mathbf{b}_{ms}.^4 \tag{27}$$

Let $\mathbf{X}_{k\times k}^{*}$ be the symmetric matrix from the $k(k+1)/2$-dimensional vector $\mathbf{x}_{k(k+1)/2}$, then

$$\mathbf{A}_{d\times k}^{\mathrm{T}}\mathbf{A}_{d\times k} = \mathbf{X}_{k\times k}^{*}. \tag{28}$$

Assume $\mathbf{X}_{k\times k}^{*}$ has rank t, then it can be factorized by the SVD decomposition as

$$\mathbf{X}_{k\times k}^{*} = \mathbf{U}_{k\times t}^{*}(\mathbf{\Sigma}_{t\times t}^{*})^2(\mathbf{U}_{k\times t}^{*})^{\mathrm{T}}. \tag{29}$$

Hence, by (28) we have

$$\mathbf{A}_{d\times k} = \mathbf{R}_{t\times t}\mathbf{\Sigma}_{t\times t}^{*}(\mathbf{U}_{k\times t}^{*})^{\mathrm{T}}, \tag{30}$$

where $\mathbf{R}_{t\times t}$ is an orthogonal matrix. Substituting Eq. (30) into Eq. (24) gives

$$\mathbf{Y}_{t\times m} = \mathbf{R}_{t\times t}\mathbf{\Sigma}_{t\times t}^{*}(\mathbf{U}_{k\times t}^{*})^{\mathrm{T}}\mathbf{Q}_{k\times m}. \qquad \square$$

6. Conclusions

The LLE is considered an effective algorithm for dimensionality reduction. In this paper, we obtained the some of its key properties: (1) for a given high-dimensional data set, there always exists a linear mapping such that all the constraint conditions in the LLE can be satisfied. The implication of the existence of such a linear mapping is that the LLE cannot guarantee a one-to-one mapping from the high-dimensional space to the low-dimensional space; (2) if the LLE is required to globally preserve distance, it must be a PCA mapping; (3) for a given high-dimensional data set, there always exists a local distance-preserving LLE. The results in this paper can bring some new insights into a better understanding of the LLE.

References

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

[2] X. He, S. Yan, Y. Hu, H.-J. Zhang, Learning a locality preserving subspace for visual recognition, in: International Conference on Computer Vision, 2003, pp. 385–393.

[3] C. Zhang, J. Wang, N. Zhao, D. Zhang, Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction, Pattern Recognition 37 (2) (2004) 325–336.

[4] A.M. Elgammal, C.-S. Lee, Separating style and content on a nonlinear manifold, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2004), 2004, pp. 478–485.

[5] N. Mekuz, C. Bauckhage, J.K. Tsotsos, Face recognition with weighted locally linear embedding, in: The Second Canadian Conference on Computer and Robot Vision, 2005, pp. 290–296.

[4] $\mathbf{B}^{+}$ is the Moore–Penrose inverse of matrix $\mathbf{B}$.

About the Author—FUCHAO WU is a professor at Institute of Automation, Chinese Academy of Sciences. His research interests are now in computer vision, which include 3D reconstruction, active vision, and image based modeling and rendering.

About the Author—ZHANYI HU is a professor at Institute of Automation, Chinese Academy of Sciences. He received his Ph.D. degree (Docteur d'Etat) in computer vision from University of Liege, Belgium, in 1993. His research interests include camera calibration and 3D reconstruction, active vision, geometric primitive extraction, vision guided robot navigation, and image-based modeling and rendering.