The Globally Linear Embedding Algorithm

Jieyun Xia
School of computer information,
Guangdong Polytechnic college,
Guangzhou, 510520,China

Shuaibin Lian*
School of Information Science and Technology,
Sun Yat-sen University,
Guangzhou, 510006,China
*Corresponding author, e-mail:shuai_lian@qq.com

***Abstract*— LLE is a very effective non-linear dimension reduction algorithm and widely explored in machine learning, pattern recognition, data mining and etc. 'Locally linear, Globally non-linear' has always been regarded as the features and advantages of LLE. However, the theoretical derivation presented in this paper shows that when the size of neighborhood is larger than the dimension of the space in which the data are presented, LLE is no longer 'global nonlinear' and almost has the same effect as PCA in dimensionality reduction. At present, a lot of literatures on LLE verify their results on Swiss Roll, Punctured Sphere, Twin Peaks, etc. These manifolds are presented in the three-dimensional Euclidean space and the size of neighborhood is always larger than three to prevent too small to be effective. But in these cases, LLE cannot play its advantage of nonlinearity.**

Keywords-Locally Linear Embedding;Dimension Reduction ; Globally Linear Embedding

I. INTRODUCTION

Dimensionality reduction is an efficient method to handle the complicated high dimensional data commonly met in our modern society and widely explored in data mining, machine learning, pattern recognition and etc. Manifold learning is considered as an ascendant method for nonlinear dimensionality reduction and becomes more and more attractive. Currently the efficient algorithms for manifold learning include the isometric mapping (Isomap)[1][2], locally linear embedding (LLE)[3][4], Laplacian Eigenmaps (LE)[5], Hessian LLE (HLLE)[6] and local tangent space alignment (LTSA)[7], etc. Among all of these methods, LLE LLE is computationally simpler and can give useful results on a broader range of manifolds[4]. Its ability to deal with large amounts of high dimensional data and its non-iterative way of finding the embeddings make it more and more attractive to researchers. The original LLE has some intrinsic drawbacks, such as sensitive to size of neighborhood and regularization parameters, ill-conditioned eigenproblems and etc. Based on these drawbacks, many extensions of LLE have been proposed recently years, reviewed in ref[8].LLE and its extensions have been used to solve various problems in machine learning[9][10][11].

The most attractive properties of LLE is 'Locally linear, Globally non-linear'[4] . Because of this, LLE has been considered as a nonlinear method of dimensionality reduction. However, the theoretical derivation presented in this paper shows that when k > n, where k is the number of neighbors of each data point and n is the dimension of the space in which the data are presented, LLE is no longer global nonlinear and degenerates into a linear method of dimensionality reduction. In this paper, for the sake of distinction, LLE is called GLE (Global Linear Embedding) when k > n.

II. LOCALLY LINEAR EMBEDDING(LLE)

Given a data matrix $X_{n\times m}=[x_1,x_2,....,x_m]$,where, x_i represent the data points in the space R^n. LEE consists of three problems and their solutions.

A. *the first problem of LLE*

For each data point x_i ,find its k neighbors, $x_{i1},x_{i2},...,x_{ik}$, there are many ways to find the neighbors[12][13][14].

B. *the second problem of LLE*

For each data point x_i and its k neighbors, $x_{i1},x_{i2},...,x_{ik}$, find k coefficients $\Omega_i=[w_{i1},w_{i2},...,w_{ik}]$ to minimize the following objective function:

$$\arg\min\left\|x_i-\sum_{j=1}^{k}w_{ij}x_{ij}\right\|^2 \tag{1}$$

subject to $\sum_{j=1}^{k}w_{ij}=1$,this is what the 'Locally Linear' means.

C. *the third problem of LLE*

Find a matrix $Y_{d\times n}$ to minimize the following objective function: $\arg\min\left\|Y_{d\times n}W_{n\times n}\right\|^2$,where $d\ll m$.The column vectors of $Y_{d\times n}$ are the data after dimensionality reduction.

III. GLOBALLY LINEAR EMBEDDING (GLE)

A. *the optimal solution to second problem of LLE*

The objective function of LLE can be rewritten as ,

$$\left\|x_i-\sum_{j=1}^{k}w_{ij}x_{ij}\right\|^2=\left\|\sum_{j=1}^{k}w_{ij}(x_{ij}-x_i)\right\|^2=\left\|G_i\Omega_i\right\|^2 \tag{2}$$

where $G_i=[x_{i1}-x_i,x_{i2}-x_i,...,x_{ik}-x_i]$, $\Omega_i=[w_{i1},w_{i2},...,w_{ik}]$,where G_i is $n\times k$

© the authors

The optimal vector Ω_i to equation (2) satisfy that $\|G_i\Omega_i\|^2=0$, $\sum_{j=1}^{k}\Omega_i=1$ By SVD decomposition, G_i can be expressed as: $G_i=U_i\sum_i(V_i)^T$,

- where $\sum_i$ is a diagonal matrix whose elements are the square roots of the nonzero eigenvalues of the matrix $G_i(G_i)^T$, $r=rank(G_i(G_i)^T)\le\min(n,k)$;
- U_i are the orthonormal eigenvectors of $G_i(G_i)^T$ corresponding to the nonzero eigenvalues.
- V_i are the orthonormal eigenvectors of $(G_i)^T G_i$ corresponding to the nonzero eigenvalues.

So when $k>n$,this means that the matrix $(G_i)^T G_i$ has $k-r$ orthonormal eigenvectors corresponding to the zero eigenvalue.

So if let $V^i_{k\times(k-r)}$ be a matrix whose column vectors are the orthonormal eigenvectors corresponding to the zero eigenvalues of $(G_i)^T G_i$, then $\left(V^i_{k\times(k-r)}\right)^T V^i_{k\times(k-r)}=0$

Now let $\Omega_i=\dfrac{V^i_{k\times(k-r)}A_{k-r}}{\Gamma_k^T V^i_{k\times(k-r)}A_{k-r}}$,where $\Gamma_k^T=[1,1,\ldots 1]$, $A_{k-r}\in R^{k-r}$

Then, $\|G_i\Omega_i\|^2=\left\|U_i\sum_i(V_i)^T\dfrac{V^i_{k\times(k-r)}A_{k-r}}{\Gamma_k^T V^i_{k\times(k-r)}A_{k-r}}\right\|^2=0$ due to $\Gamma_k^T\Omega_i$ =1, this means $\Omega_i=\dfrac{V^i_{k\times(k-r)}A_{k-r}}{\Gamma_k^T V^i_{k\times(k-r)}A_{k-r}}$ is the optimal solution to the second problem of LLE. Note that $A_{k-r}\in R^{k-r}$ provide a freedom of degree for Ω_i

The above derivation shows that, when *k* > *n*, no matter how to choose neighbors for each data point, the optimal solution to the second problem of LLE can be found. Furthermore, in this case, $X_{m\times n}W_{n\times n}=0$.

B. the optimal solution to the third problem of LLE

A matrix $Y_{d\times n}$ is said to be the optimal solution of LLE if and only if $\|Y_{d\times n}W_{n\times n}\|^2=0$, $Y_{d\times n}Y^T_{d\times n}=I$

By SVD decomposition, $X_{m\times n}$ can be expressed as $X_{m\times n}=U_{m\times r}\sum_{r\times r}V^T_{r\times n}$,

- where $r=rank(X_{m\times n}X^T_{m\times n})=rank(X^T_{m\times n}X_{m\times n})$
- $\sum_{r\times r}$ is a diagonal matrix whose elements are the square roots of nonzero eigenvalues of $X_{m\times n}X^T_{m\times n}$. Note that the nonzero eigenvalues of $X_{m\times n}X^T_{m\times n}$ are the same as those of $X^T_{m\times n}X_{m\times n}$.
- The column vectors of $U_{m\times r}$ are the orthonormal eigenvectors of $X_{m\times n}X^T_{m\times n}$ corresponding to the nonzero eigenvalues.
- The column vectors of $V_{r\times n}$ are the orthonormal eigenvectors of $X^T_{m\times n}X_{m\times n}$ corresponding to the nonzero eigenvalues.

As previously demonstrated in section 3.1, when *k* > *n*, $X_{m\times n}W_{n\times n}=0$,which lead to the following derivation: $X_{m\times n}W_{n\times n}=U_{m\times r}\sum_{r\times r}V^T_{r\times n}W_{m\times m}=0$,so we have, $V^T_{r\times n}W_{m\times m}=\sum^{-1}_{r\times r}U^T_{m\times r}0_{m\times n}=0$

Now let $Y_{d\times n}=P_{d\times r}V^T_{r\times n}$, where $P_{d\times r}P^T_{d\times r}=I$. This means that $Y_{d\times n}=P_{d\times r}V^T_{r\times n}$ is the optimal solution to the third problem of LLE. Furthermore, since $X_{m\times n}=U_{m\times r}\sum_{r\times r}V^T_{r\times n}$, so we have $V^T_{r\times n}=\sum^{-1}_{r\times r}U^T_{m\times r}X_{m\times n}$, the optimal solution $Y_{d\times n}$ can be repressed as $Y_{d\times n}=P_{d\times r}V^T_{r\times n}=P_{d\times r}\sum^{-1}_{r\times r}U^T_{m\times r}X_{m\times n}=C_{d\times m}X_{m\times n}$,

where $C_{d\times m}=P_{d\times r}\sum^{-1}_{r\times r}U^T_{m\times r}$

C. the relationship of LLE and PCA

From our derivation, we can see that the optimal solution $Y_{d\times n}$ can be expressed as the linear combination of $X_{m\times n}$.

So in this case LLE turned to be a linear dimension reduction method as PCA(Principle Components Analysis) which is commonly used of linear dimension reduction method.

IV. EXPERIMETAL RESULTS

Figure 1－6 and 7－12 shows the experimental results of GLE and PCA on some classical manifolds such as Swill Roll, Gaussion, Toroidal Helix, TwinPeaks, Conner Plane and Puncture Sphere. As can be seen from these Figures, GLE and PCA almost have the same effect. Note that PCA is a linear method of dimensionality reduction. The experimental results show from another perspective that when *k* > *n*, LLE is not a global non-linear, but a global linear method of dimension reduction.

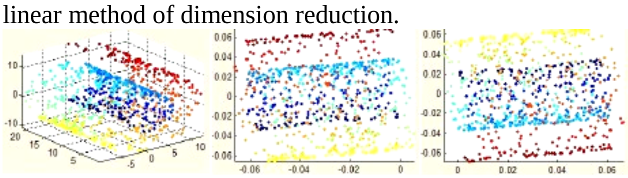

Figure1: original manifold(left),GLE(middle),PCA(right) on Swiss Roll with n=3,m=800,k=8

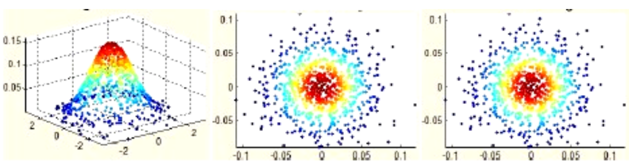

Figure2:original manifold(left),GLE(middle),PCA(right) on Gaussion with n=3,m=800,k=8

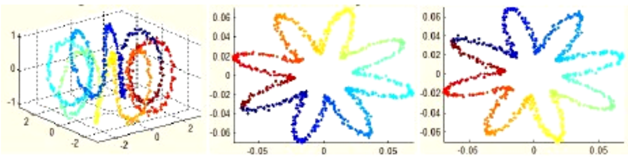

Figure3:original manifold(left),GLE(middle),PCA(right) on Toroidal Helix, with n=3,m=800,k=8

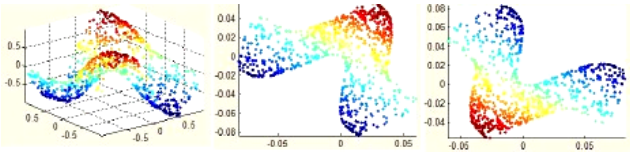

Figure4:riginal manifold(left),GLE(middle),PCA(right) on Twin Peaks, with n=3,m=800,k=8

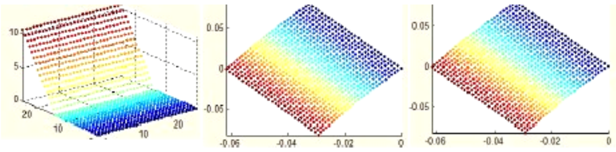

Figure5: original manifold(left),GLE(middle),PCA(right) on Conner Plane, with n=3,m=800,k=8

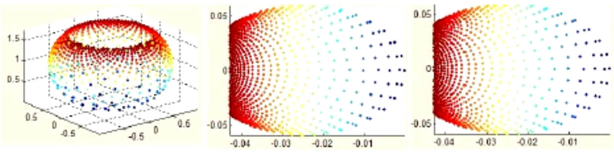

Figure6: original manifold(left),GLE(middle),PCA(right) on Puncture Sphere, with n=3,m=800,k=8

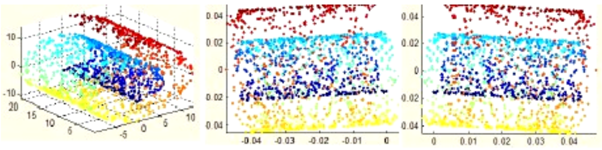

Figure7: original manifold(left),GLE(middle),PCA(right) on Swiss Roll with n=3,m=1500,k=12

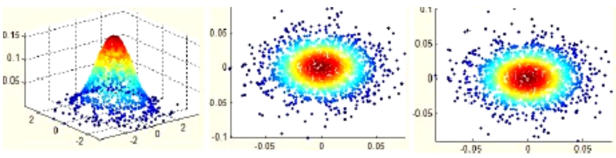

Figure8: original manifold(left),GLE(middle),PCA(right) on Gaussion with n=3,m=1500,k=12

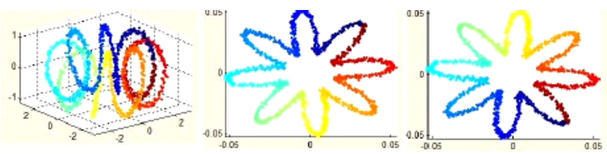

Figure9: original manifold(left),GLE(middle),PCA(right) on Toroidal Helix with n=3,m=1500,k=12

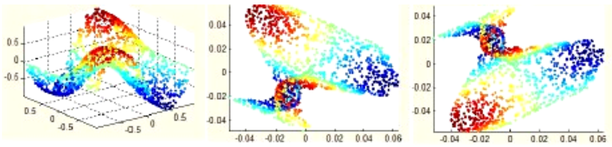

Figure10: original manifold(left),GLE(middle),PCA(right) on Twin Peaks with n=3,m=1500,k=12

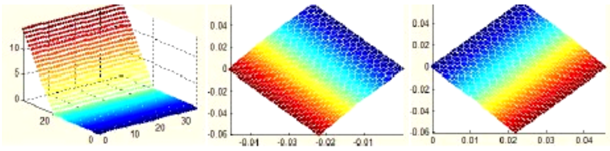

Figure11: original manifold(left),GLE(middle),PCA(right) on Conner Plane with n=3,m=1500,k=12

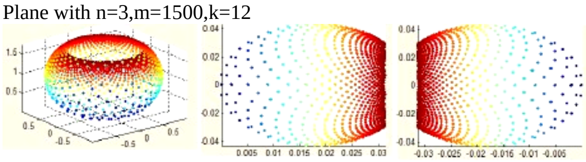

Figure12: original manifold(left),GLE(middle),PCA(right) on Puncture Sphere with n=3,m=1500,k=12

From these experimental results, we can safely come to a conclusion that in this case, when *k*>*n*, LLE is globally linear dimension reduction method.

V. CONCLUSION

Dimension reduction is a very effective method to deal with high dimensional data and widely explored in pattern recognition, data mining, machine learning and etc. LLE has been regarded as the classical non-linear dimension reduction algorithm because of this good performances on real data and artificial data. 'Globally non-linear, Locally linear' is the main feature of LLE. Duo to these characteristics LLE has been regarded as the non-linear dimension reduction methods. But from the derivation presented in this paper, we can clearly come to some conclusions as follows:

- when *k*>*n*, we can have $X_{m\times n}W_{n\times n}=0$,LLE turn out to be GLE. The sufficient and necessary condition for $X_{m\times n}W_{n\times n}\neq 0$ is that there is at least one point $x_i \notin \left\{ \sum_{j=1}^{k} w_{ij}x_{ij} \,\middle|\, w_{ij}\in R,\ j=1,2,...,k;\ \sum_{j=1}^{k} w_{ij}=1 \right\}$, this case will not happen, when *k*>*n.*

- In order to be visual, a lot of literatures on LLE verify their results on the data taken from the manifolds such as Swiss Roll, Punctured Sphere, Twin Peaks, etc. All of these manifolds are presented in the three dimensional Euclidean space *R*3(*n* = 3) and the number of neighbors is always larger than 3, i.e., *k* > *n*, otherwise, the size of neighborhood would be too small to be effective. As demonstrated in this paper,

© the authors

when $k > n$, LLE degenerates into a linear method of dimensionality reduction and is no longer ”local linear and global nonlinear”. Therefore, the manifolds such as Swiss Roll, Punctured Sphere, Twin Peaks, etc, may not be suitable for the verification of LLE.

REFERENCES

[1]. Joshua B Tenenbaum, Vin De Silva, J. C. Langford, A global geometric framework for nonlinear dimension reduction, Science 290 (2000) 2319–2323.

[2]. M. Balasubramanian, E. L. Schwartz, J. B. Tenenbaum, V. de Silva, J. C. Langford, The Isomap algorithm and topological stability, *Science* 4 January 2002, 295: 7

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

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

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

[6]. D. Donoho, C. Grimes, Hessian Eigenmaps: new tools for nonlinear dimensionality reduction, Proceedings of National Academy of Science(2003) 5591-5596.

[7]. Z. Zhang, H. Zha, Principal manifolds and nonlinear dimensionality reduction via tangent space alignment, SIAM Journal of Scientific Computing 26 (2005) 313-338.

[8]. Jing chen and Yan Liu. Locally Linear Embedding: a suvery. Artif Intell Rev (2011) 36:29–48

[9]. WangY, WuY. Complete neighborhood preserving embedding for face recognition .Pattern Recognition, 2010, 43(3) :1008–101

[10]. Yan Y, Zhang Y .Discriminant projection embedding for face and palmprint recognition. Neurocomputing ,2008,71(16–18):3534–3543

[11]. Ying HP, Andrew Teoh BJ, Wong EK. Neighbourhood discriminant embedding in face recognition. IEICE Electron Express . (2008),5(24):1036–1041

[12]. Pan Y, Ge SS, Al Mamun A . Weighted locally linear embedding for dimension reduction. Pattern Recognition,2009, 42(5):798–811

[13]. GuihuaW, Lijun J, JunW. Kernel relative transformation with applications to enhancing locally linear embedding. IJCNN,2008, pp 3401–3406

[14]. Valencia-Aguirre J, Álvarez-Mesa A, Daza-Santacoloma G, Castellanos-Domínguez G. Automatic choice of the number of nearest neighbors in locally linear embedding CIARP,2009, pp 77–84

© the authors