Natl. Acad. Sci. Lett. (May–June 2017) 40(3):189–196
DOI 10.1007/s40009-017-0536-7

SHORT COMMUNICATION

Modified Locally Linear Embedding with Affine Transformation

Kanika Mehta[1] · Gajendra Tyagi[1] · Aashish Rao[1] · Pardeep Kumar[1] · Durg Singh Chauhan[2]

Received: 5 February 2015 / Revised: 12 April 2016 / Accepted: 3 January 2017 / Published online: 21 January 2017

Abstract Dimensional reduction is a primary way to analyze and work with complex and large amount of multidimensional data by avoiding the effect of curse of dimensionality. This problem of constructing low dimensional embedding gains importance in number of fields like artificial intelligence, image processing, geographical research and lot more. In this paper, we introduce a modified locally linear embedding, an unsupervised learning algorithm that computes low dimensional data from complex high dimensional data using affine transformation and neighborhood preserving embedding. Unlike novel locally linear embedding, our method is affine invariant where each point is being represented by an affine combination of its neighboring points. At the end, we conduct the experiment to evaluate our proposed method and compare its performance with existing methods. Results show that our proposed method is unaffected by affine transformation, specifically shear while existing methods fail to produce correct results in case of shear.

✉ Pardeep Kumar
pardeepkumarkhokhar@gmail.com

Kanika Mehta
er.kanmehta@gmail.com

Gajendra Tyagi
gajendratyagi007@gmail.com

Aashish Rao
rao01aashish@gmail.com

Durg Singh Chauhan
pdschauhan@gmail.com

[1] Department of Computer Science and Engineering, Jaypee University of Information Technology, Solan, India

[2] GLA University, Mathura, India

Keywords Locally linear embedding · Affine transformations · Unsupervised learning

Introduction

In the field of data mining, machine learning, information processing, data compression, and artificial intelligence, preprocessing is an important step to refine the data for any further processing. The main purpose of this refinement is to derive a representation that is more understandable and meaningful to apply operations such as classification, interpolation, outlier detection, or visualization [1]. It also makes it easy to interpret the information from given data. As most of the real world data like global climate patterns, human gene distributions, spectrogram of speeches consists of large number of dimensions that imposes challenge to traditional preprocessing techniques. Dimension reduction is a technique to handle such high dimensional data. It is one of the main goals of unsupervised learning that is extracting hidden information from huge and complex data with large number of dimensions without any prior knowledge of data. However, till now, there are number of techniques that have been proposed to overcome the effect of this "curse of dimensionality" [2].

Dimensionality reduction is a well known problem from many years [3]. Several methods have been developed and implemented on non linear multidimensional data. Initially two methods principle component analysis [4] and multidimensional scaling [5] based on eigenvectors involve the modeling of linear variability in multidimensional data. In PCA [4], linear projections are computed on the basis of greatest variance calculated from the top eigenvectors of data covariance matrix. MDS [5] works on pair wise distance matrix and computes the lower dimensional embedding. The result of both the methods is equivalent if the distance matrix corresponds to Euclidean distance. Limitation of applicability on only linear multidimensional data of these two methods introduced the concept of locally linear embedding (LLE) which is also applicable to non-linear multidimensional data. In year 2000, Roweis and Saul [6], presented the new method especially for nonlinear multidimensional data and named it as locally linear embedding. Local geometric structure is constructed by the method which is invariant to translations and orthogonal transformations in a neighborhood of each data point and maps the data points to a lower dimension in which local geometries are also remain preserved. In LLE, each data point Xi having dimensions D is used to find out the optimized weights w_{ij} to the neighbor (j). Then these weights are used to construct the data points Yi in the lower dimensions d. This is how, LLE is able to preserve the local geometries. Tenenbaum et al. [7] in year 2000, presented the method ISOMAP, for computing a low dimensional embedding of a set of high dimensional data points. It is basically an extension to MDS in which, geodesic distances are incorporated. MDS works on pair wise distances to the data points which is generally measured using Euclidean distance, while ISOMAP consider the geodesic distance generated by a neighborhood graph embedded in the classical scaling. Geodesic distance is the sum of weights along the shortest path between two nodes. Demerit of PCA and MDS which is their inability to discover non liner degree of freedom that underlies complex natural observations such as images of a face, handwriting is removed by ISOMAP. In year 2001 [8], Mikhail Belkin and Partha Niyogi proposed an algorithm based on laplacian Beltrami operator. Their approach for representing high dimensional data is based on correspondence between the graph laplacian Beltrami operator and connections to heat equations. Computationally, it was an efficient approach for dimensionality reduction. In this, the approximation of manifold is done by the adjacency graph computed from data points. The weights are chosen by the heat kernel which is made useful by the Laplace–Beltrami operator in the heat equation. One more key point of this algorithm is it's insensitivity to the outliers and noise. This algorithm cannot embed out of sample points, but techniques based on Reproducing kernel Hilbert space regularization exist for adding this capability. Donoho and Grimes [9] in 2003 came out with Hessian method to recover the underlying parameterization of scattered data (mi) lying on a manifold M embedded in high-dimensional Euclidean space. It was a modification in locally linear embedding and Laplacian eigenmap framework, where they substituted a quadratic form based on Laplacian to a form based on hessain. Hessain LLE is a sparse matrix based technique which gives better results than LLE, but the main drawback of hessian LLE is that it's computational complexity and therefore it is not suitable for heavy sampled manifolds. Zhenyue and Hongyuan [10] in year 2004 came out with an approach which is based on a parameterized manifold which is used to form a sample of unorganized data points. It works on the fact that all of the tangent hyper planes will be aligned if manifold is correctly unfolded. It's a two way algorithm which can efficiently learn a nonlinear embedding into lower dimension coordinates from higher dimension data and vice versa. Coifman et al. [11] in year 2005 utilized diffusion process to find the meaningful embedding of datasets in a lower dimensional Euclidian space whose coordinates are computed by using the eigen values and eigenvectors of diffusion operator on data. It basically lies in the family of nonlinear dimensionality reduction whose main focus is to find underlying manifold of the data has been sampled from. This method is basically using the eigen functions of a Markov matrix which is defining a random walk on the data to obtain new descriptions of data sets. The non-linearity and preserving of local structures are the two main advantages of this algorithm over the classical dimensionality reduction algorithms. Zhenyue and Wang [12] in year 2007 presented Modified LLE which addresses the problem of regularization i.e. when the number of neighbors is greater than the number of input dimensions. The matrix defining each local neighbor rank deficient. MLLE solved this problem by using multiple weight vectors in each neighborhood. The algorithm is divided into three stages (1) finding nearest neighbors, (2) weight matrix construction, (3) Partial eigen value decomposition (same as LLE). It requires N neighbors greater than n dimensions ($N > n$). Zhenyue et al. [13] in year 2012 proposed Adaptive Manifold Learning that addresses two main issues. First is the adaptive neighborhood size selection through neighborhood expansion and contraction. Second is the adaptive bias reduction in embedding by weighting local affine errors in the embedding of the manifold which is specially designed for LTSA by modifying the minimization model in original LTSA. The advantage of this approach is that it works well for noisy data sets. But it is very difficult to find out a good embedding with a noisy data set that is sampled from a manifold with variable curvatures and the noises are relatively large. Yu et al. [14] in year 2014 presented a novel supervised learning algorithm for representing a high dimensional data to low dimension. It is a supervised algorithm [13] in which class labels of the data points are needed to be known for the classification purpose. This classification is done to get the similar data points together and separate dissimilar data points. This modification in LLE [4] overcame the problem of topological distortion of low dimension data with uneven distribution in high dimensional space. This paper also introduces the concept of area based Chi square discretization that effectively discretizes continuous data to discrete data in low dimensional space. The only disadvantage with this approach is that the class label of the data points should be known and it is not invariant to affine transformations such as shear. Guangbin et al. [15] in 2014 came out with an improvement over the LLE in which traditional Euclidean distance is replaced by homogenization distance to construct the weight matrix which reduces the impact of neighboring points in the dimension reduction. The homogenization distance is calculated as given in the equation.

$$Dij = \frac{||xi - xj||}{\sqrt{M(i)M(j)}}$$

Here ||xi − xj|| is the Euclidean distance and M(i) and M(j) respectively represents the average distance of xi and xj to other points. The only purpose of using homogenization distance is to get the narrow distance in relatively sparse area and increased distance in the dense area. This paper doesn't address the problem of affine transformations. It is just reducing the impact of neighboring points in the embedding formation. Zhihui et al. [16]. extended the LLE method to sparse cases by proposing a unified sparse learning framework and L1-norm learning. Robustness in the classification or feature extraction can be enhanced by L1-norm sparse learning. Experimental results proved the performance of Sparse linear embedding (SLE) and Sparse kernel Embedding (SKE) for dimensionality reduction in terms of preserving the local geometric structures and orthogonality and thus robustness. But the paper didn't talk about affine transformation. Xianglei et al. [17] in April 2015 presented the fusion of different manifold learning methods for dimensionality reduction. Authors worked on the believe that each method is based on some different geometric foundation and thus they learn different partial information from the original geometric embedding. So they just combined and reformulated the laplacian eignmaps, LLE and Hessian LLE to get the optimized manifold structure in low dimensional coordinate space. Dayong and Juan [18] in 2015 proposed a supervised learning algorithm for dimension reduction in speech recognition system. The algorithm uses locality preserving projection method of multi kernel and after pre-processing methods like speech framing and feature extraction, a 50*12 dimension matrix is obtained. With reduced computation, the algorithm achieved good recognition effect but again the paper does not show the effect of affine transformation. In this paper, a new improved method that is modified locally linear embedding for affine transformations [19] is introduced which is locally affine invariant and can handle more complex and natural transformations of an object. Affine transformations are transformations that preserves lines and parallelism and are composed of linear transformation (shear, scaling or rotation) and a translation. The affine transformations have a property that they preserve the co linearity relation between the points, that is point which lie on same line continue to be collinear after the transformation. In a high dimension space affine transformation locally looks like rotation plus translation which leads to local isometry but for non-neighbors it acts like scaling resulting in distortion of global geometry. Figure 1 shows the embedding formed by different methods discussed above like MDS, PCA, ISOMAP, LLE, HLLE, Laplacian, Diffusion map and LTSA. The topmost left plot represents the original Swiss roll dataset [20] in 3D space (x, y and z coordinates) with 800 points and 8 nearest neighbors while all other plots represents the embedding formed by different specified methods in 2D space (x and y). Also the time taken by each method is specified in seconds (s).

The basic LLE is considered as a foundation in the field of nonlinear dimensional reduction. It is based on simple intuition of geometry that computes a low dimensional embedding from high dimensional space keeping the intrinsic correlation of the original data. The LLE algorithm as the name suggests, reconstructs the data points locally where only the neighbors contribute to each reconstruction that is confined to linear subspace.

The first step is to define k nearest neighbors using Euclidean distance for each data points. Using cost function we compute reconstruction error E(W) as :

$$E(W) = \sum_i \left|\left| Xi - \sum_j WijXj \right|\right| 2\ldots \tag{1}$$

The weight w_{ij} represents the role of the jth data point in the ith reconstruction. These weights are computed using the cost function in Eq. (1) that is minimized and subjected to sparseness and invariant constraints. The sparseness constraint states that each data point X_i is reconstructed only from its neighbors, making these data points invariant to scaling and rotation imposing $W_{ij} = 0$ if X_j does not belong to this set. The invariance constraint states that rows of the weight matrix sum to one that is $\sum_j W_{ij} = 1$ making the data point invariant to translation. Finally, at the end each higher dimension input X_i is mapped on to a lower dimension output Y_i by selecting d dimensional coordinates of Y_i to minimize embedded cost function Φ(Y):

$$\Phi(Y) = \sum_i \left|\left| Yi - \sum_j WijYj \right|\right| 2\ldots \tag{2}$$

The weight w_{ij} represents the role of the jth data point in the ith reconstruction and Y_i is the data point reconstructed only from its neighbors Y_j. Figure 2 shows the three steps involved in the basic LLE. The above discussed method fails when working with affine transformations that include shear [1]. So to overcome this problem we propose our modified locally linear embedding with affine transformations.

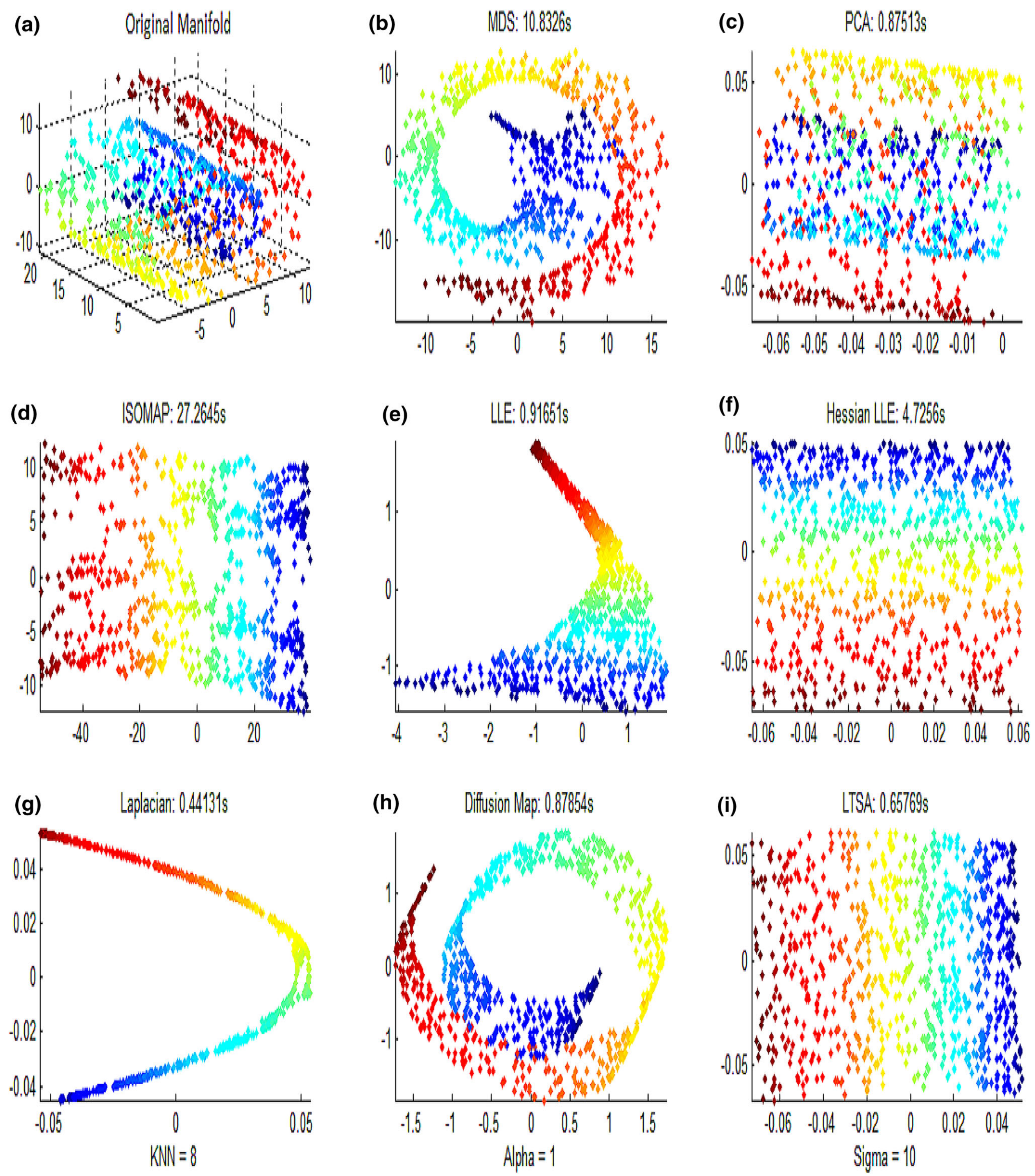

Fig. 1 Embedding formed by different dimensionality reduction algorithms. **a** Original embedding of the manifold, **b** embedding formed by multidimensional scaling, **c** principle component analysis, **d** embedding formed by ISOMAP, **e** locally linear embedding, **f** hessian LLE, **g** laplacian, **h** diffusion map, **i** local tangent space alignment

To explain our modified locally linear embedding with affine transformation (LLEWAF), Let P = {p1, p2, p3, …, pn} be a given dataset of N samples, and D is the dimension of the dataset. We first apply maximum likelihood estimation (MLE) method to estimate the intrinsic dimension (ID) d from P in Step 1. Unlike LLE that

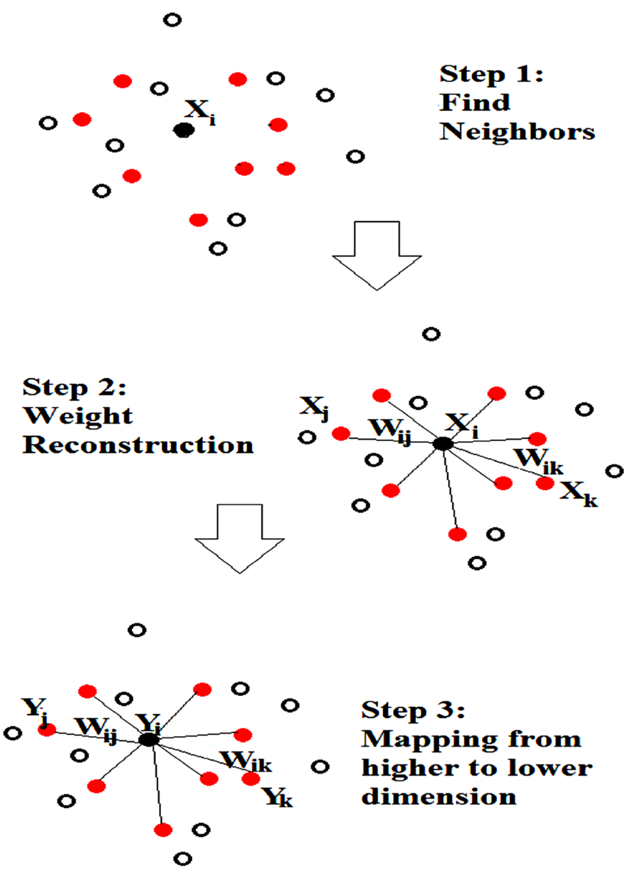

Fig. 2 Three steps involved in basic locally linear embedding to form embedding in lower dimension space

Table 1 Parameter settings for the experiment

Technique	Parameter settings
PCA	None
ISOMAP	K = 12
LLE	K = 12
Hessian LLE	K = 12
Laplacian	K = 12
LTSA	K = 12
MLLE	K = 12

considers each point to be represented by the convex combination, we have considered each point to be represented by the affine combination of its neighboring points. Then we apply clustering algorithm to find k nearest neighbors with a constraint that number of neighbors must at least be three and should be non collinear as the affine hull of a set of three points not on one line is the plane going through them. Using affine combination of neighbors of pi we can represent pi by:

$$Pi = \sum_{pj \in Np} Wijpj\ldots \quad (3)$$

Such that W represents an $n_t * n_t$ weight matrix of affine combination coefficients for all pi. The weight matrix W is subjected to two constraints. First is a sparseness constraint that states that each data point pi is reconstructed only from its neighbors, describing only the local geometric properties of each point. This enforces $W_{ij} = 0$ if p_j does not belong to N_p. Second is invariance constraint such that each row of the weight matrix sum to one: $\sum_j W_{ij} = 1$, making the representation invariant to global translation. To minimize the error of each points affine combination we use constrained least square problem as follows:

$$\arg\min \left\| pi - \sum_j Wijpj \right\| 0, 1, 2, \ldots n \quad for\ i = 1, \ldots np\ldots \quad (4)$$

Such that $\sum_j W_{ij} = 1$. The error function (2) is affine invariant:

$$= \arg\min \left\| pi - \sum_j Wijpj \right\| 2$$

$$= \arg\min \left\| Api - \sum_j Wij\ Apj \right\| 2$$

$$= \arg\min \left\| (pi + t) - \sum_j Wij\ (pij + t) \right\| 2\ldots \quad (5)$$

where A and t denote an arbitrary 2*2 affine transformation matrix and an arbitrary 2*1 translation vector. Once the reconstruction weights are computed, modified NLLE maps P to Y in a lower dimensional space R_d according to (4):

$$\arg\min yi \sum_{i \in Np} \left\| yi - \sum_{j \in k} WijYj \right\| 2, s.t\ \mathrm{Yt\ Y} = 1\ldots \quad (6)$$

Given below is the algorithm for the proposed method

Algorithm : Modified LLE with Affine Transformation

Input: Dataset P of N samples with dimension D
k: number of nearest neighbors at each sample point pi
Output: Embedded result Y with the intrinsic dimension d

1) Estimate the global intrinsic dimension d of the dataset P using the MLE method.
2) Find k nearest neighbors of each pi such that k is greater than equal to 3 and neighbors are non collinear.
3) Calculate local reconstruction weights.
4) Apply affine transformation.
5) Map the dataset P to Y in a lower dimensional space R_d.

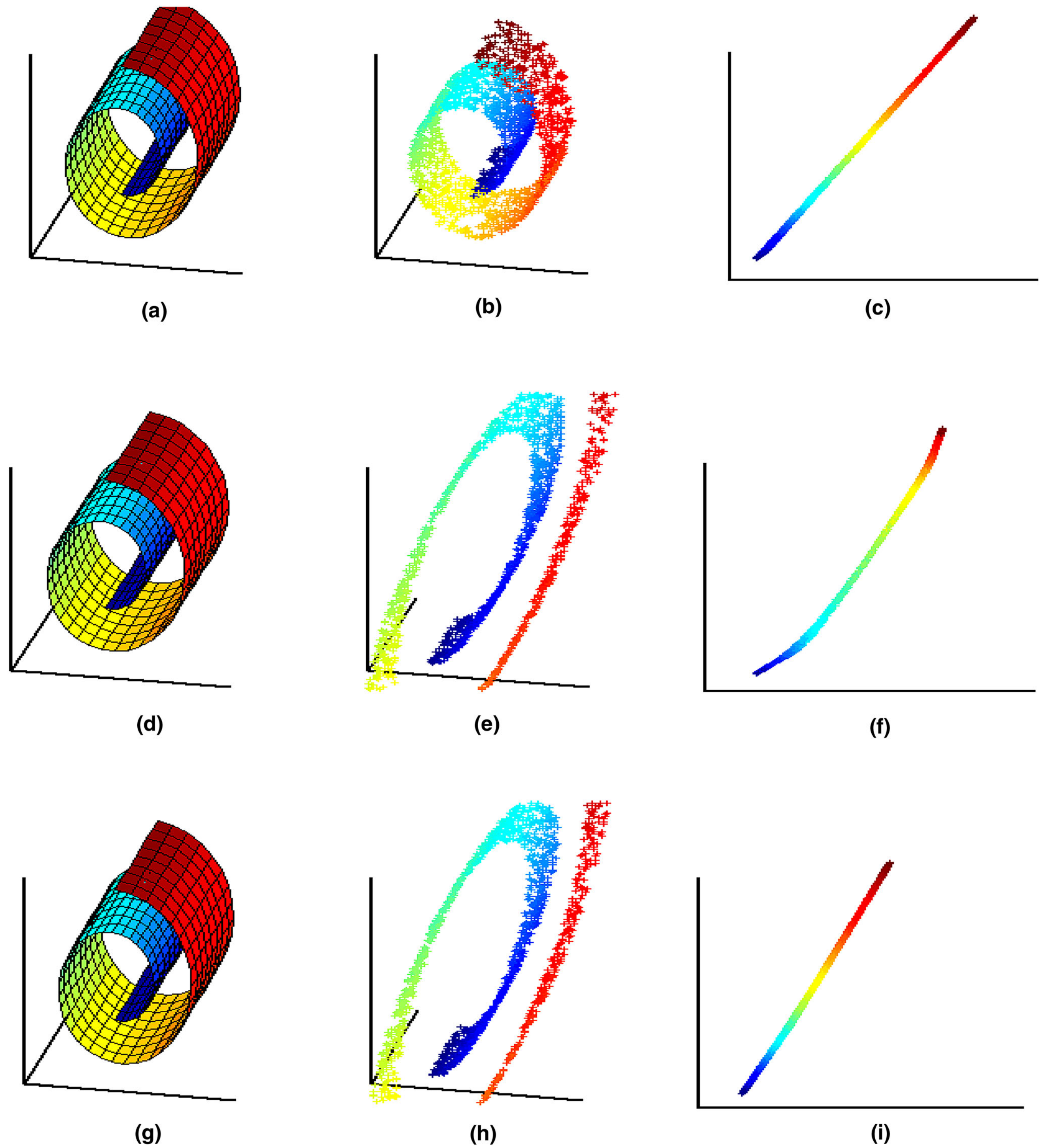

Fig. 3 Embedding formed by local linear embedding and modified LLEWAF. **a** Original embedding of the manifold, **b** Scattered plot of sampled data, **c** Embedding formed using LLE **e** Scattered plot of sheared sampled data, **f** distorted plot formed by LLE, **i** embedding formed using modified LLE with affine transformation

Implementation and Result: In order to evaluate our proposed method in a real-world situation, we conducted experiments on a Lenovo S210 with 2-GB RAM and Intel(R) Pentium(R) CPU 2127U @1.90 GHz, on Windows 8(64-bit Operating System, x-64based processor).We have taken Swiss roll dataset (P) with input consisting of N = 2000 data points taken from the original manifold and k = 12 neighbors per data point as shown in

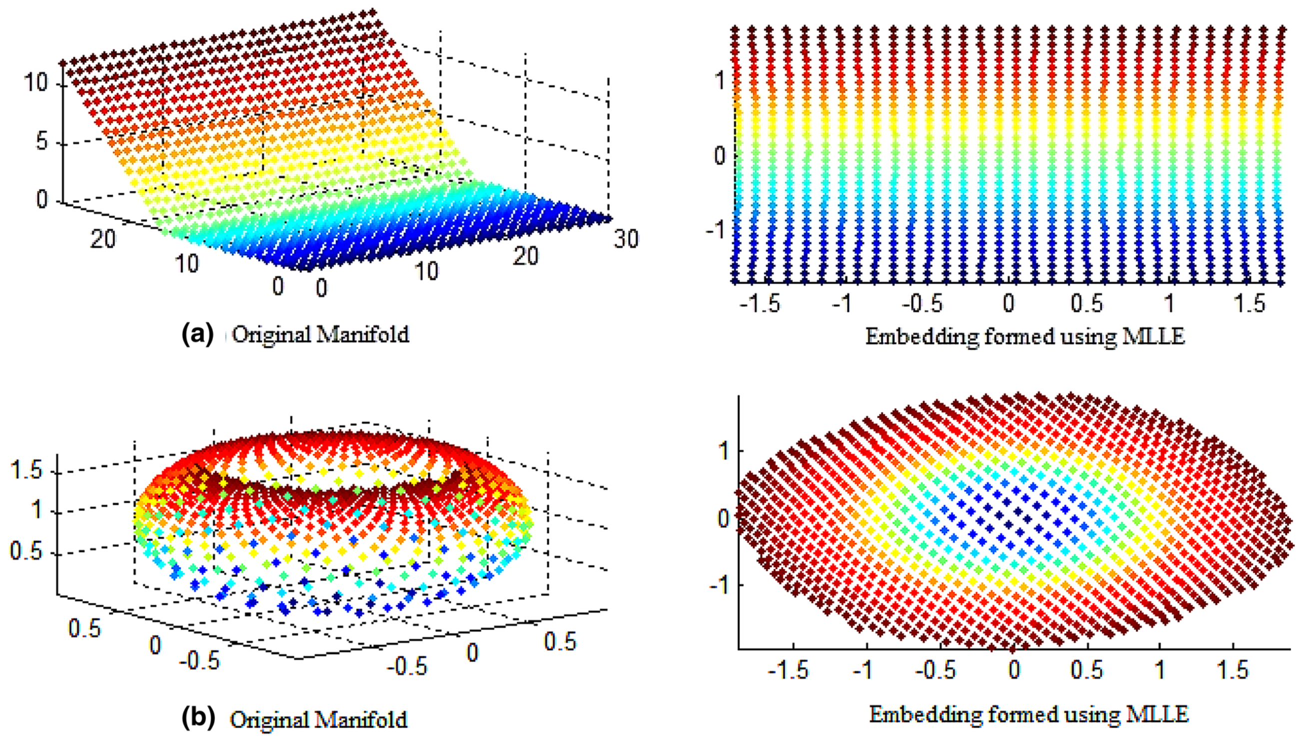

Fig. 4 Embedding formed by modified local linear embedding on **a** corner planes, **b** punctured sphere

Table 2 Generalization error on artificial data using linear discriminant classifier

Dataset	PCA (%)	ISOMAP (%)	LLE (%)	Hessian LLE (%)	Laplacian (%)	Diffusion Map (%)	LTSA (%)	MLLE (%)
Swiss roll	55.64	1.86	10.98	1.17	23.87	14.64	1.13	1.05
Corner planes	34.65	15.42	12.43	6.44	13.48	11.34	15.67	12.43
Punctured sphere	25.63	9.89	17.87	8.77	18.53	18.74	25.45	10.82
Average	38.64	9.06	13.76	5.46	18.63	14.91	14.08	**8.10**

Table 1. We applied Locally Linear Embedding (LLE) and Modified Locally Linear Embedding with Affine Transformation (MLLEAF) in MATLAB to analyze the effect of shear transformation on final embedding formed in lower dimension. We took Swiss roll dataset with d = 3 (dimensions), k = 1 (neighbors) and N = 2000 (data points) shown in Fig. 3a. Using this dataset we reconstructed a scatter plot (Fig. 3b). We applied LLE algorithm to this scattered plot and obtained a new embedding in 2D space (Fig. 3c). In order to show the effect of shear, we took the sampled data P again and multiplied it with affine transformation matrix A (also called the shear matrix) i.e. P*A. Now we applied LLE on this new sheared sampled data and found that this new embedding in 2D is distorted and different from what was expected (Fig. 3f). To overcome this problem of distortion due to shear, we applied our algorithm. We took the original sampled data and applied affine transformation to it. We then obtained a scattered plot of sheared sampled data and applied our proposed MLLEWAT. It was found that the new embedding formed is similar to LLE without affine transformation (Fig. 3i). This is because our proposed algorithm (MLLEWAF) considered each point to be represented as affine combination of its neighboring points, unlike LLE which considered convex combination of each point. We also checked our algorithm on two other datasets shown in Fig. 4 which are Corner planes and Punctured sphere with 1000 points. We can see that the final embedding is without affine transformation similar to LLE. Table 2 shows the generalization errors of linear discriminant classifiers trained using artificial datasets like Swiss roll. The leftmost column shows the datasets and topmost row contains all the techniques used for dimensionality reduction. We can observe that the proposed modified LLE is not best for every dataset, but on an average, it has the minimum generalization error.

In this paper, we have proposed a modified locally linear embedding with affine transformations, a non-supervised dimensional reduction method which overcomes the drawback of locally linear embedding (LLE) and its extensions. In this method we considered the affine combination of points instead of convex combination to show the effects of affine transformations. Unlike LLE and NLLE, we considered only those neighbors who are non-collinear because of which our proposed method is not affected by affine transformations such as shear. First, we reconstructed weights using our modified k nearest neighbor algorithm, applied affine transformation and then mapped the given high dimension data to low dimension embedding. Our experimental results on real data demonstrate that MLLE produces correct and faithful embedding with affine transformation like shear.

References

1. Lawrence KS, Sam TR (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. J Mach Learn Res 4:119–155
2. John AL, Michel V (2007) Nonlinear dimensionality reduction. Springer, New York
3. Shuicheng Y, Dong X, Benyu Z, Hong JZ, Qiang Y, Stephen L (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29(1):40–51
4. Jolliffe IT (1986) Principal component analysis. Springer, New York
5. Cox T, Cox M (1994) Multidimensional scaling. Chapman and Hall, London
6. Sam TR, Lawrence KS (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326. doi:10.1126/science.290.5500.2323
7. Tenenbaum JB, Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323
8. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and representation. Neural Comput 15(6):1373–1396
9. Donoho DL, Grimes G (2003) Hessian eigenmaps: locally linear embedding techniques for high dimensional data. Proc Natl Acad Sci 100(10):5591–5596
10. Zhen-yue Z, Hong-yuan Z (2004) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. J Shanghai Univ 8(4):406–424
11. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, Zucker SW (2005) Geometric diffusion as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc Natl Acad Sci 102(21):7426–7431
12. Zhenyue Z, Wang J (2007) MLLE: modified locally linear embedding using multiple weights. In: Proceedings of advances in neural information processing systems, vol 19, pp 1593–1600
13. Zhenyue Z, Jing W, Hongyuan Z (2012) Adaptive manifold learning. IEEE Trans Pattern Anal Mach Intell 34(2):253–265
14. Yu S, Heng Q, Keqiu L, Yingwei J, Deqin Y, Shusheng G (2015) An effective discretization method for disposing high-dimensional data. Inf Sci 270:73–91
15. Guangbin W, Jun L, Yilin H, Qinyi C (2015) Fault diagnosis of supervision and homogenization distance based on local linear embedding algorithm. Math Probl Eng 2015:981598. doi:10.1155/2015/981598
16. Zhihui L, Wai KW, Yong X, Jian Y, David Z (2015) Approximate orthogonal sparse embedding for dimensionality reduction. IEEE Trans Neural Netw Learn Syst 27(4):723–735. doi:10.1109/TNNLS.2015.2422994
17. Xianglei X, Kejun W, Zhuowen L, Yu Z, Sidan D (2015) Fusion of local manifold learning methods. IEEE Signal Process Lett 22(4):395–399
18. Dayong Z, Juan W (2015) Speech recognition using locality preserving projection based on multi kernel learning supervision. In: International symposium on computers and informatics, pp 1508–1516
19. Modenov PS, Parkhomenko AS (2014) Euclidean and affine transformations: geometric transformations. Academic Press, New York
20. Tenenbaum JB (1998) Mapping a manifold of perceptual observations. In: Proceedings of advances in neural information processing systems, vol 10, pp 682–688