

ISOLLE: Locally Linear Embedding with Geodesic Distance

Claudio Varini^{1,2}, Andreas Degenhard², and Tim Nattkemper¹

¹ Applied Neuroinformatics Group, Faculty of Technology,
University of Bielefeld, Bielefeld, Germany

{cvarini, tnattkem}@techfak.uni-bielefeld.de

² Condensed Matter Theory Group, Faculty of Physics,
University of Bielefeld, Bielefeld, Germany

adegenha@physik.uni-bielefeld.de

Abstract. Locally Linear Embedding (LLE) has recently been proposed as a method for dimensional reduction of high-dimensional nonlinear data sets. In LLE each data point is reconstructed from a linear combination of its n nearest neighbors, which are typically found using the Euclidean Distance. We propose an extension of LLE which consists in performing the search for the neighbors with respect to the geodesic distance (ISOLLE). In this study we show that the usage of this metric can lead to a more accurate preservation of the data structure. The proposed approach is validated on both real-world and synthetic data.

1 Introduction

The analysis of complex high-dimensional data structures is essential in many real-world applications, including medical high resolution time-series data. Algorithms for dimensional data reduction are particularly useful for discerning the information contained in a high-dimensional data structure.

In recent years several methods for the analysis of nonlinear data sets have been proposed, including Locally Linear Embedding (LLE) [1]. LLE has already been successfully applied to many problems, including face recognition [2], prediction of membrane protein types [3] and the analysis of micro array data [4]. The algorithm assumes linearity in the local area centered on each data point. Each area is mathematically characterized by a set of coefficients (weights) which correlate the particular data point with its n nearest neighbors. The aggregation of all areas can be intuitively thought as an assemblage of linear patches which approximates the nonlinear data structure. The high-dimensional data is then projected into a lower-dimensional space while preserving the coefficients between neighboring data points.

The number of neighbors n strongly influences the accuracy of the linear approximation of nonlinear data. Specifically, the smaller n , the smaller the area, the more faithful is the linear approximation. However, if these areas are disjoint, LLE can fail to detect the global data structure [5]. Disjoint areas can be obtained especially when the data is sparse or spread among multiple clusters.

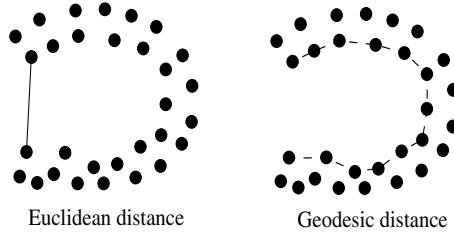


Fig. 1. The short circuit induced by Euclidean distance is shown on the left. In case the number of neighbors n is set to a relative high value, the two points in figure can be treated as neighbors, although they are on the opposite parts of the horseshoe. This may cause LLE to fail to detect the real global structure of the data. On the right are shown the benefits of the geodesic distance. In this case the two points are not neighbors, as they are faraway according to the geodesic distance.

To address this problem, in [6] it is proposed to search for the $n/2$ nearest and $n/2$ furthest neighbors of each data point. Another approach is given in [7], where the authors suggest to connect the disjoint manifold or interpolating the embeddings of some samples.

In general, for larger values of n the linear areas are more likely to overlap. The number of neighbors n therefore needs to be sufficiently high to satisfy this condition. On the other hand, as the neighbors search is typically conducted using the Euclidean distance, this may lead a data point to have neighbors which are instead very distant as one considers the intrinsic geometry of the data. More intuitively, one can imagine this fact as a short circuit (see Fig. 1). The presence of short circuits is undesirable, as they can cause LLE to misinterpret the actual data structure.

To address the above outlined problems occurring to LLE when employed with a high number of neighbors, we propose the usage of LLE with geodesic distance (ISOLLE). More specifically, the n nearest neighbors are searched with respect to the geodesic distance. This metric has already been employed in other methods for nonlinear dimensional data reduction such as Isomap [8], Curvilinear Distance Analysis [9] and Self-organizing Maps [10]. The geodesic distance between two data points can be intuitively thought as their distance along the contour of an object (see Fig. 1 right). For example let us consider the distance between Paris and New York. Their geodesic distance is the distance along the curvature of the Earth. Their Euclidean distance instead is the length of the straight line connecting the two cities which is below the level of the ground. Points faraway from each other, as measured by the geodesic distance, may appear deceptively close in the high-dimensional input space as measured by the Euclidean distance.

In this work we demonstrate that the employment of the geodesic distance can lower the probability to create short circuits during the neighbors search, thereby allowing for a more accurate dimensional reduction. Our approach to investigate the performances of ISOLLE as compared to conventional LLE is

basically twofold. Firstly, we perform the analysis on synthetic data, namely a three-dimensional swissroll which was also used in [1] and [8]. By this phantom data set we illustrate the difference between both techniques. Secondly, we analyze a complex, medical real-world data set acquired using dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). DCE-MRI involves the repeated imaging of a region of interest, in our case the female breast with tumor lesions, after the administration of a contrast agent, yielding a high-dimensional spatio-temporal data structure.

Both data sets are reduced to two dimensions using different values of the number of neighbors n . The dimensional reduction of the swissroll is evaluated qualitatively, while the analysis of the tumor data set requires a statistical approach because of the complexity of the data. Specifically for this purpose we consider the percentage of nearest points in the original space that are preserved as nearest neighbors in the dimensional reduced space, and the stress induced by the dimensional reduction. In addition, in the final part the running times of LLE and ISOLLE are compared.

2 Locally Linear Embedding (LLE)

The LLE algorithm is based on three steps involving standard methods of linear algebra. Its input comprises N D -dimensional vectors $\{\mathbf{X}_i\}$. The first step consists in searching for the n nearest neighbors of each data point.

Once the neighbors are determined, by minimizing the following error function (step 2)

$$\Psi(W) = \sum_{i=1}^N \left| \mathbf{X}_i - \sum_{j=1}^n W_{ij} \mathbf{X}_j \right|^2 \quad (1)$$

subject to the constraint $\sum_{j=1}^n W_{ij} = 1$, one obtains the weights $\{W_{ij}\}$ that best allow to reconstruct each data point from its neighbors. With the above constraints, Eq. (1) can be simplified to a linear system and the weights can be computed in closed form as follows: given a particular data point \mathbf{X}_i with n -nearest neighbors \mathbf{X}_j and reconstruction weights W_j that sum to one, we can write the reconstruction error as

$$\Psi(W) = \sum_{i=1}^N \left| \mathbf{X}_i - \sum_{j=1}^n W_j \mathbf{X}_j \right|^2 = \sum_{jk} W_j W_k C_{jk}. \quad (2)$$

In the second identity, the term

$$C_{jk} = (\mathbf{X}_i - \mathbf{X}_j) \cdot (\mathbf{X}_i - \mathbf{X}_k) \quad (3)$$

is the local covariance matrix. The weights which minimize the error function of Eq. (1) are given by:

$$W_j = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}}, \quad l, m \in \{1, \dots, n\}. \quad (4)$$

In some cases, for example if the number of neighbors is greater than the input dimension ($n > D$), it arises that the matrix C is singular or nearly singular and the solution of Eq. (2) is not unique. In this case the matrix C must be conditioned by adding a small multiple of the identity matrix [11]:

$$C_{ij} \leftarrow C_{ij} + \delta_{ij}\Gamma \quad (5)$$

where Γ is defined as

$$\Gamma = \frac{\text{Tr}(C)}{n}\Delta^2. \quad (6)$$

The term Δ is a correction parameter set by the user and its value must be much smaller than 1.

The third and last step of the LLE algorithm consists in mapping each data point \mathbf{X}_i to a low dimensional vector \mathbf{Y}_i , such that the following embedding error function is minimized:

$$\Phi(Y) = \sum_{i=1}^N |\mathbf{Y}_i - \sum_{j=1}^n W_{ij} \mathbf{Y}_j|^2 \quad (7)$$

under the conditions $\frac{1}{N} \sum_{i=1}^N \mathbf{Y}_i \mathbf{Y}_i^T = I$ and $\sum_{i=1}^N \mathbf{Y}_i = 0$, which provide a unique solution. Note that the weights are kept constant in order to preserve the local neighborhood of each data point. The most straightforward method for computing the M -dimensional coordinates is to find the bottom $M + 1$ eigenvectors of the sparse matrix

$$S = (I - W)^T(I - W). \quad (8)$$

These eigenvectors are associated with the $M + 1$ smallest eigenvalues of S . The bottom eigenvector is related to the smallest eigenvalue whose value is closest to zero. This eigenvector is the unit vector with all equal components and is discarded.

3 The ISOLLE Algorithm

The ISOLLE algorithm differs from LLE only in the first step, i.e. the neighbors search. More specifically, ISOLLE computes the n nearest neighbors of each data point according to the geodesic distance. For this purpose we employ a small variation of Dijkstra's algorithm [12]. Given a graph, this algorithm computes the shortest paths from a particular node to all remaining nodes. In our case we restrict the computation to the n shortest paths.

In practice, the process of finding the geodesic neighbors is composed of two phases. The first phase consists in constructing a weighted graph G over the data set where neighboring data points are connected. In principle, any similarity measure d_E can be adopted to determine neighboring relations, and probably the Euclidean distance is the most common choice. Two points are neighbors if are closer than a fixed distance ϵ (ϵ -graph), or one is the K nearest point of the

other (K -graph). These relations between neighbors are represented by edges of weights $d_E(\mathbf{X}_i, \mathbf{X}_j)$ [8].

In the second phase the n nearest neighbors of each data point are found according to the geodesic distance computed by Dijkstra's algorithm. This algorithm begins at a specific node (source vertex) and extends outward within the graph until all the vertices have been reached (in our case only the n nearest nodes). Dijkstra's algorithm creates labels associated with vertices. These labels represent the distance (cost) from the source vertex to that particular vertex. Within the graph, there exists two kinds of labels: temporary and permanent. The temporary labels are given to vertices that have not been reached. The value given to these temporary labels can vary. Permanent labels are given to vertices that have been reached and their distance (cost) to the source vertex is known. The value given to these labels is the distance (cost) of that vertex to the source vertex. For any given vertex, there must be a permanent label or a temporary label, but not both. An animated example of Dijkstra's algorithm can be seen at [13]. Both steps of the neighboring search are detailed in the following:

Construct the neighborhood graph: define the graph G over all data points by connecting points \mathbf{X}_i and \mathbf{X}_j if (as measured by $d_E(\mathbf{X}_i, \mathbf{X}_j)$) they are closer than ϵ , or if \mathbf{X}_i is one of the K nearest neighbors of \mathbf{X}_j . Set edge lengths equal to $d_E(\mathbf{X}_i, \mathbf{X}_j)$.

Compute n nearest points with Dijkstra's algorithm: given a graph $G=(V,E)$ where V is a set of vertices and E a set of edges, Dijkstra algorithm keeps two sets of vertices:

S –the set of vertices whose shortest paths from the source vertex have already been determined. These vertices have a permanent label

V-S –the remaining vertices. These have a temporary label

The other data structures needed are:

\mathbf{X}_0 –initial beginning vertex (source vertex)

N –number of vertices in G

\mathbf{D} –array of estimates of shortest path to \mathbf{X}_0 .

The basic mode of operation of Dijkstra's algorithm is:

1 $S=\{\mathbf{X}_0\}$

2 For $i=1$ to N
 $\mathbf{D}[i]=E[\mathbf{X}_0,i]$

3 For $i=1$ to $N-1$

 Choose a vertex w in $V-S$ such that $\mathbf{D}[w]$ is minimum and add it to S

 For each vertex v in $V-S$

$\mathbf{D}[v]=\min(\mathbf{D}[v],\mathbf{D}[w]+E[w,v])$

The construction of graph G requires a further parameter (ϵ or K) to be set by the user. In [8] it is pointed out the scale-invariant parameter K is typically easier to set than ϵ , but may yield misleading results when the local dimensionality varies across the data set. A sensible way to set this parameter can be to choose the minimal value such that all the pairwise geodesic distances are finite.

4 Data Sets

The performances of ISOLLE and LLE are tested on two data sets whose numbers of points are displayed in table 1. The first data set is a three-dimensional synthetic distributions, namely a swissroll (Fig. 2(a)). The second is a real-world data set comprising the signal intensity values obtained by dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) on female breast with tumor.

The DCE-MRI technique consists in acquiring a sequence of images (six in our case) from a region of interest (the female breast in our case), whose movement is carefully restricted, over time in order to monitor the dynamic of a previous injected contrast agent within the tissue. As a result, tissue types with higher level of vascularity have enhanced values of signal intensity, proportionally to the amount of absorbed contrast agent. After the acquisition of the images, a time-series of intensity values is correlated with each voxel (see Fig. 2(b)). As benign and malignant tumor tissues are expected to differ in the level of vascularization, the respective contrast characteristics are expected to exhibit different behaviors.

The tumor data set in this work comprises the time-series relative to six benign and six malignant cancerous lesions which were labeled and pathologically analyzed by an expert physician. The discrimination between benign and malignant lesions in DCE-MRI is a particularly challenging and delicate problem in light of the relatively high rate of false positive cases characterizing this imaging technique, as published in the literature [14]. In this study the time-series associated with the voxel of each tumor is treated as a data point in a six-dimensional signal space. It is therefore interesting to project this six-dimensional space in two dimensions in order to visualize how benign and malignant data differs from each other.

5 Method for Comparing ISOLLE and LLE

The difference between LLE and ISOLLE are illustrated by considering the three-dimensional swissroll. At first we visualize the neighbors graphs obtained by LLE

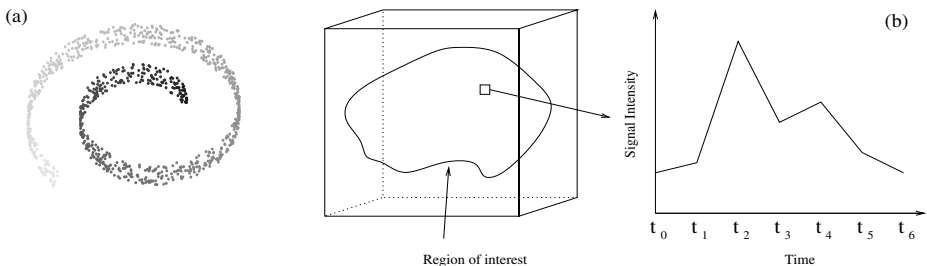


Fig. 2. (a) Three-dimensional swissroll data set. (b) In DCE-MRI, a time-series of MR signal intensity values is associated with each voxel.

Table 1. Data sets investigated in this study

Data set	Number of points	Dimension
Swissroll	1000	3
DCE-MRI breast tumor data	2449	6

and ISOLLE with different values of n in order to highlight the advantages of the geodesic distance. Each graph is obtained by connecting each data point with its n neighbors by an edge (note that this is not the graph G used to compute the geodesic distances and described in section 3). This allows us to check for the possible presence of short circuits induced by the neighbors search. The effects of these short circuits are then qualitatively evaluated by visualizing the respective two-dimensional projections.

The evaluation of the dimensional reduction of the tumor data set requires a statistical analysis, as the output can not be predicted a priori because of the complexity and multi-dimensional nature of the data, and consequently it is not possible to visually evaluate the accuracy of the low-dimensional projection.

The quality of the tumor data embeddings is estimated by means of two numerical quantities, namely neighborhood preservation (NP) and stress (ST). The first quantity is given by the average percentage of neighbors which are preserved after the dimensional reduction. It is defined as

$$\text{NP} = \frac{1}{V} \sum_{i=1}^V p_t(\mathbf{X}_i) \quad (9)$$

where $p_t(\mathbf{X}_i)$ is the percentage of the t -nearest neighbors of point \mathbf{X}_i in the original space which are preserved in the low-dimensional space. For example, if only 25% of its t -nearest neighbors are preserved in the embedding, then $p_t(\mathbf{X}_i)$ will equal 0.25. In this work we use $t = 5$. A high value of NP (close to 1) denotes a good preservation of the local relations between data points in the low-dimensional space.

Stress reflects the preservation of the global structure of the original data set in the embedding. More specifically, it quantifies the overall deviation (i. e. the extent to which they differ) between the distances in the original and embedded space [15]. Let \mathbf{X}_i and \mathbf{X}_j be two data points; their distance in the original and in the embedding space are indicated by $d(\mathbf{X}_i, \mathbf{X}_j)$ and $\delta(\mathbf{X}_i, \mathbf{X}_j)$, respectively. Stress is typically defined in terms of variance as

$$\text{ST} = \frac{\sum_{\mathbf{X}_i, \mathbf{X}_j} (\delta(\mathbf{X}_i, \mathbf{X}_j) - d(\mathbf{X}_i, \mathbf{X}_j))^2}{\sum_{\mathbf{X}_i, \mathbf{X}_j} d(\mathbf{X}_i, \mathbf{X}_j)^2}. \quad (10)$$

Prior to the computation of the value of stress, both the original and embedded coordinates are scaled to $[0,1]$ in order to allow for a correct comparison between different embeddings. Low values of stress (close to 0) reflect a good preservation of the original pairwise distances in the low dimensional space.

6 Experiments

Graph G is computed by setting ϵ to the minimal possible value such that all the pairwise distances are finite. These values empirically found for each data set are: $\epsilon(\text{swissroll})=5$; $\epsilon(\text{tumor data})=90$.

The two data sets are reduced to two dimensions by LLE and ISOLLE with the number of neighbors n varying between 5 and 40.

7 Results and Discussion

In Fig. 3 one can see the neighbors graphs of the swissroll. It is obvious that already with $n = 15$ LLE with Euclidean distance meets some short circuit effects in the neighbors search. With $n = 40$ the number of short circuits increases noticeably. By contrast, the graphs relative to ISOLLE do not present short circuit effects, even when the number of neighbors n equals 40. This shows that the usage of the geodesic distance can drastically reduce the number of short circuits.

Possible effects of these short circuits on the two-dimensional projection of the swissroll data set can be seen in Fig. 4. Here it is clear that LLE fails to preserve the global structure of the data with $n = 15$ and in particular $n = 40$, as in both cases the darkest points are mapped close to brighter points. On the contrary, ISOLLE can correctly unfold the swissroll in all three cases, and the structure of the data is clearly preserved. In particular, the ISOLLE projection is also accurate with $n = 40$, while the respective LLE projection results completely incorrect.

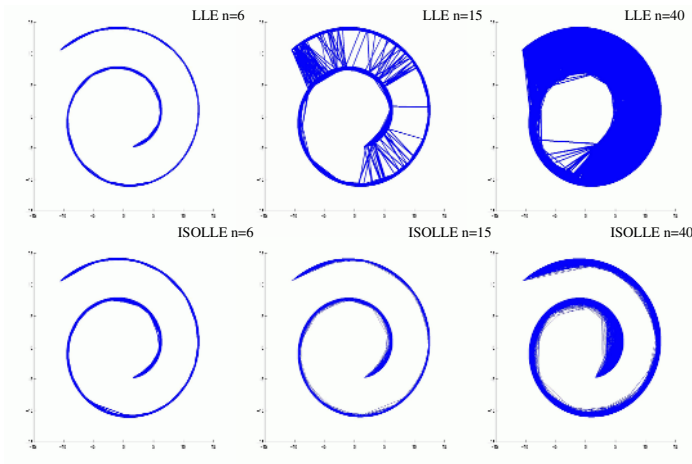


Fig. 3. Neighbors graphs of the swissroll data set. In the LLE graph with $n = 15$ there are already short circuits. Their number considerably increases with $n = 40$. Conversely, in all the ISOLLE graphs there are no short circuits.

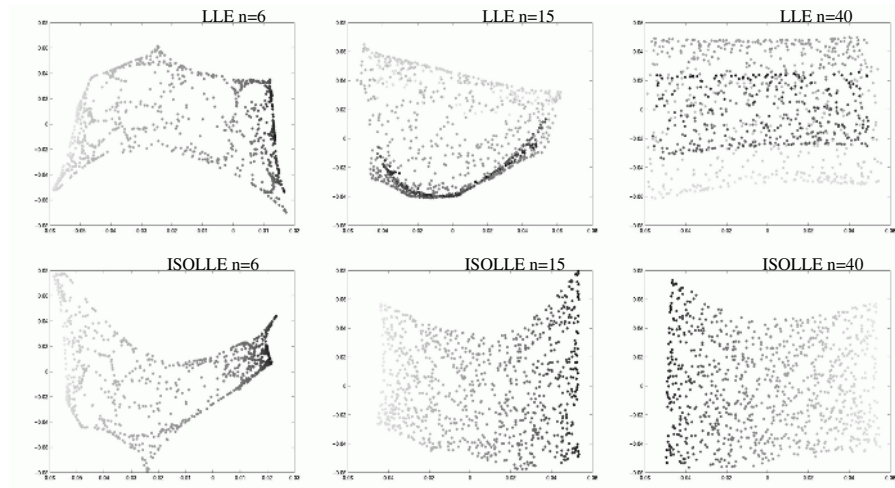


Fig. 4. Two-dimensional reductions of the swissroll data set. While LLE fails to preserve the structure of the swissroll with $n \geq 15$, ISOLLE yields a good projection of the data in all cases.

The evaluation of dimensional reduction of the tumor data set is conducted by taking into account the neighborhood preservation and stress measures. Their average values with the respective variances computed with respect to n comprised between 5 and 40 are displayed in table 2. The projections by ISOLLE result better with respect to both quantities. Indeed, the average ST value is lower than the one by LLE, suggesting that ISOLLE better preserves the metric of the tumor data. The higher value of the average NP by ISOLLE gives evidence that this algorithm also leads to a better preservation of the topology of the data. Two scatter plots of the DCE-MRI breast data embeddings obtained by LLE and ISOLLE with $n = 20$ are shown in Fig. 5. Interestingly, the benign cluster in the projection by ISOLLE appears more localized and compact than in the projection by LLE. Moreover, benign and malignant data overlap more in the projection by LLE. This indicates that ISOLLE can better separate benign from malignant data and this is of considerable value from the medical point of view. In addition, the compactness of the benign cluster in the ISOLLE projection shows that benign tumors are rather homogeneous, while the malignant ones are more heterogeneous, in agreement with the clinical experience of physicians [16].

Table 2. Average and variance values of stress (ST) and neighborhood preservation(NP) computed for the tumor data set

ST(LLE)	ST(ISOLLE)	NP(LLE)	NP(ISOLLE)
0.454±0.034	0.337 ± 0.025	0.081±0.001	0.115 ± 0.002

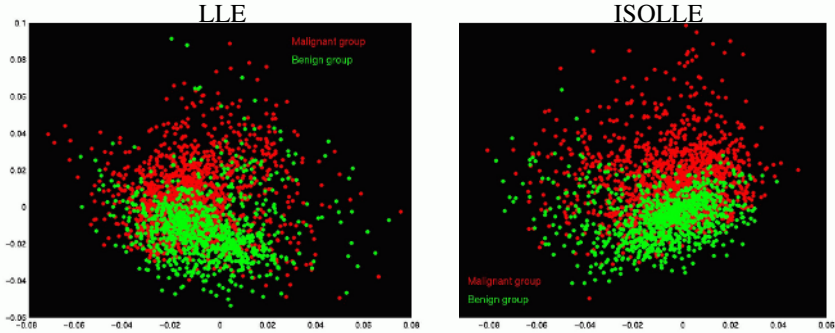


Fig. 5. Two scatter plots of the two-dimensional embeddings of the DCE-MRI breast data set obtained by LLE and ISOLLE. In both cases n equals 20. Note that the benign and malignant clusters overlap much less in the ISOLLE embedding. In particular, here the benign cluster is more compact and localized.

Table 3. Table of the running times in seconds

n	Swissroll		DCE-MRI	
	LLE	ISOLLE	LLE	ISOLLE
10	0.20	2.66	1.26	16.55
20	0.23	6.32	1.39	39.24
30	0.27	11.25	1.62	69.31
40	0.30	17.29	1.75	106.37

Finally, we compare the performances of LLE and ISOLLE in terms of running time. Both algorithms were run with different n on a Pentium IV 2.8 GHz. The respective values of running times are shown in table 3. The ISOLLE algorithm involves a larger computation time and the divergence of speed becomes more marked as n increases. The higher computational time of ISOLLE is somewhat expected as the algorithm requires a further step as compared to LLE, i. e. the construction of the neighborhood graph over all data points.

In general, the usage of ISOLLE should be preferred to LLE in particular when n needs to be relatively high (for example in case of sparse or clustered data) and in turn short circuits are more likely to occur. One way to determine if a certain data set requires a relatively high value of n is to perform an analysis of the smallest eigenvalues of matrix S from eq. (8). Specifically, in standard conditions matrix S has only one eigenvalue close to 0. However, if n is so small that the linear areas are disjoint, then matrix S will have more than one close-to-zero eigenvalue [5]. Therefore, the minimum n for which S has only one eigenvalue close to 0 can be taken into account in order to evaluate which algorithm is more suited for the data analysis.

8 Conclusions

In this study we propose a new approach to the neighbor search in LLE based on the geodesic distance. Its usage can reduce the number of short circuits considerably, thereby improving the preservation of the data structure. We show this by investigating the neighbors graphs obtained by LLE and ISOLLE on a synthetic three-dimensional swissroll. The ISOLLE graphs do not exhibit short circuits, even when the number of neighbors is high. By contrast, the standard neighbors search with Euclidean distance in LLE causes many short circuits. As a consequence, ISOLLE can detect the intrinsic two-dimensional structure of the swissroll with both small and large values of the number of neighbors n . Conversely, LLE fails to unfold the swissroll with $n \geq 15$.

Regarding the dimensional reduction of the tumor data set, our results clearly show that ISOLLE significantly outperforms LLE in terms of both stress and neighborhood preservation. In addition, ISOLLE appears to better distinguish between benign and malignant lesions.

Experiments concerning the running times revealed that ISOLLE is slower than LLE and this becomes more noticeable as n increases.

In conclusion, ISOLLE exhibits a superior ability to project the investigated data sets into a two-dimensional space while preserving the original data structure but at the cost of a larger running time.

Future work will include the comparison of the performances of LLE and ISOLLE with respect to data sets having different density distributions, with particular regard to sparse data.

References

1. Roweis, S. T. and Saul, L. K. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290:2323–2326, 2000.
2. Zhang, J., Shen, H. and Zhou, Z. H. Feature Dimension Reduction for Microarray Data Analysis using Locally Linear Embedding. In *SINOBIOMETRICS*, pages 296–304, 2004.
3. Wang, M., Yang, J., Xu, Z. and Chou, K. SLLE for predicting Membrane Proteins Types. *Journal of Theoretical Biology*, 232:7–15, 2005.
4. Chao, S. and Lihui, C. Feature Dimension Reduction for Microarray Data Analysis using Locally Linear Embedding. In *3rd Asia-Pacific Bioinformatic Conference*, 2004.
5. Polito, M. and Perona, P. Grouping and Dimensionality Reduction by Locally Linear Embedding. In *Neural Information Processing Systems NIPS*, 2001.
6. Vlachos, M. et al. Non-Linear Dimensionality Reduction Techniques for Classification and Visualization. In *8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002.
7. Hadid, A. and Pietikäinen, M. Efficient Locally Linear Embeddings of Imperfect Manifolds. In *MLDM*, 2003.
8. Tenenbaum, J. B., de Silva, V. and Langford, J. C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290:2319–2322, 2000.

9. Lee, J.A., Lendasse, A., Donckers, N. and Verleysen, M. A robust nonlinear Projection Method. In *Proceedings of ESANN 2000*, pages 13–20, Bruges, Belgium, 2000.
10. Wu, Y.X. and Takatsuka, M. The geodesic self-organizing Map and its Error Analysis. In *Australian Computer Science Conference*, volume 38, pages 343–352, Newcastle, Australia, 2005.
11. Lawrence, J. B., and Roweis, S. T. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003.
12. Dijkstra, E. W. A Note on two Problems in Connection with Graphs. *Numer. Math*, 1:269–271, 1959.
13. <http://www.cs.sunysb.edu/~skiena/combinatorica/animations/dijkstra.html>.
14. Kelcz, F., Furman-Haran, E., Grobgeld, D. et al. Clinical Testing of High-Spatial-Resolution Parametric Contrast Enhanced MR Imaging of the Breast. *AJR*, 179:1485–1492, 2002.
15. Hjaltason, G. R., and Samet, H. Properties of Embeddings Methods for Similarity Searching in Metric Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):530–549, 2003.
16. Furman-Haran, E., Grobgeld, D., Kelcz, F. et al. Critical Role of Spatial Resolution in Dynamic Contrast-Enhanced Breast MRI. *Journal of Magnetic Resonance Imaging*, 13:862–867, 2001.