

Topology Preservation Measures in the Visualization of Manifold-Type Multidimensional Data

Rasa KARBAUSKAITĖ, Gintautas DZEMYDA

*Institute of Mathematics and Informatics
Akademijos 4, 08663 Vilnius, Lithuania
e-mail: {karbauskaite, dzemyda}@ktl.mii.lt*

Received: September 2008; accepted: May 2009

Abstract. Most of real-life data are not often truly high-dimensional. The data points just lie on a low-dimensional manifold embedded in a high-dimensional space. Nonlinear manifold learning methods automatically discover the low-dimensional nonlinear manifold in a high-dimensional data space and then embed the data points into a low-dimensional embedding space, preserving the underlying structure in the data. In this paper, we have used the locally linear embedding method on purpose to unravel a manifold. In order to quantitatively estimate the topology preservation of a manifold after unfolding it in a low-dimensional space, some quantitative numerical measure must be used. There are lots of different measures of topology preservation. We have investigated three measures: Spearman's rho, König's measure (KM), and mean relative rank errors (MRRE). After investigating different manifolds, it turned out that only KM and MRRE gave proper results of manifold topology preservation in all the cases. The main reason is that Spearman's rho considers distances between all the pairs of points from the analysed data set, while KM and MRRE evaluate a limited number of neighbours of each point from the analysed data set.

Keywords: dimensionality reduction, manifold learning, multidimensional data visualization, locally linear embedding, topology preservation.

1. Introduction

Data coming from the real world are often difficult to understand because of its high dimensionality. There are many methods for dimensionality reduction and its further visualization. The common goal of visualization is to represent data from a high-dimensional space in a low-dimensional projection space so as to preserve the "internal structure" of the data in the high-dimensional space as far as possible and to allow the visual insight into complex multidimensional data sets. In this paper, we concentrate on the dimensionality reduction methods that deal with specific data – manifold type multidimensional data.

Most of real-life data are multidimensional, but they are not truly high-dimensional. They are just embedded in a high-dimensional space, but can be efficiently summarized in a space of a much lower dimensionality, such as a nonlinear manifold. A manifold is

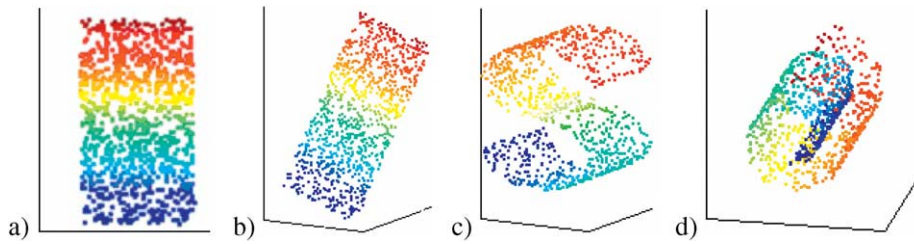


Fig. 1. A two-dimensional manifold (a), embedded in three dimensions: (b) a linear embedding (plane), (c) an S-shape, (d) “Swiss roll”.

a smooth low-dimensional surface embedded in a higher dimensional space. A simple example is given in Fig. 1. A two-dimensional manifold (Fig. 1a) is embedded in three dimensions in three different ways: a linear embedding (plane) – Fig. 1b, an S-shape – Fig. 1c, a “Swiss roll” – Fig. 1d. The S-shape manifold and “Swiss roll” can be thought of as curling a piece of rectangular paper (Fig. 1a). The analysed data sets are presented in Figs. 1b, 1c, and 1d. The aim is to transfer these data into a lower-dimensional space. In the general case, the problem is to discover the low-dimensional nonlinear manifold in a high-dimensional data space and then transfer the data points into this low-dimensional space.

Let the dimensionality of analysed data be n . High-dimensional data sets can have meaningful low-dimensional structures hidden in the observation space, i.e., the data are of a low intrinsic dimensionality $d \ll n$, in the sense of lying on or near to a smooth low-dimensional manifold. The intrinsic dimensionality of a data set is usually defined as the minimal number of parameters or latent variables necessary to describe the data.

An important point of a manifold is its topology, i.e., neighbourhood relationships between the subregions of the manifold. A manifold can be entirely characterized by giving relative or comparative proximities: a first region is close to a second one, but far from a third one.

Nonlinear manifold learning methods are topology-preserving methods. The key purpose of such methods is to preserve distances when mapping data to a low-dimensional space so that the points, close in the high-dimensional input space, be also close in the output space: it is necessary to unfold a nonlinear manifold.

A large number of nonlinear manifold learning methods has been proposed over the last decade: Locally Linear Embedding (LLE) (Roweis and Saul, 2000; Saul and Roweis, 2003), Isomap (Tenenbaum *et al.*, 2000), Laplacian Eigenmaps (LE) (Belkin and Niyogi, 2003), Hessian LLE (HLLE) (Donoho and Grimes, 2005), Local Tangent Space Analysis (LTSA) (Zhang and Zha, 2004), and others (Lee and Verleysen, 2007). While Locally Linear Embedding, Laplacian Eigenmaps, Hessian LLE, Local Tangent Space Analysis try to preserve the local geometry of the manifold, Isomap aims at preserving geometry at all scales: local and global.

This paper deals with a Locally Linear Embedding method (Roweis and Saul, 2000; Saul and Roweis, 2003). The LLE algorithm requires these parameters to be determined: the intrinsic dimensionality d and the number of the nearest neighbours k . Improper va-

lues of these parameters greatly influence the results. On one hand, a large value of the intrinsic dimensionality d amplifies noise effects, while a low value leads to overlaps in mapping results (excessively reduced; Yin *et al.*, 2007). It is noted in Levina and Bickel (2005) that if the dimensionality d is too small, important data features are “collapsed” onto the same dimensionality, and if the dimensionality is too large, the projections become noisy and, in some cases, unstable. In Karbauskaitė *et al.* (2007), it is shown that, if k is set too small, a continuous manifold can falsely be divided into disjoint sub-manifolds, and thus, the mapping does not reflect any global properties. If k is too high, a large number of the nearest neighbours causes smoothing or elimination of small-scale structures in the manifold, the mapping loses its nonlinear character and behaves like traditional Principal Component Analysis (Jolliffe, 1989).

In this paper, two-dimensional manifolds embedded in a three-dimensional space are investigated. In this case $n = 3$ and $d = 2$, i.e., we can see the embedded manifolds visually. As the structure of a manifold is known in advance, the visual impression is used in order to evaluate the topology preservation when the manifolds are unravelled. We also use several quantitative numerical measures to assess the embeddings computed by LLE. We know what embeddings must be gotten on a plane. So we could say which topology preservation measures are proper for this task. Such measures can be successfully used to estimate the low-dimensional embeddings of high-dimensional data while looking for the proper value of the parameter k .

There are a lot of different measures of topology preservation in the literature (Siegel and Castellan, 1988; Goodhill and Sejnowski, 1996; König, 2000; Tenenbaum *et al.*, 2000; Venna and Kaski, 2001; Lee and Verleysen, 2007 etc.).

Different topology preservation measures are appropriate for different applications (Goodhill and Sejnowski, 1996). Our purpose is to find and investigate such measures that were suitable to analyse the topology preservation of a manifold. In this paper, we investigate three main quantitative measures: Spearman’s rho (Siegel and Castellan, 1988), König’s measure (König, 2000), and mean relative rank errors (Lee and Verleysen, 2007). Spearman’s rho is often used for estimating the topology preservation with a view to reduce dimensionality (Bezdek and Pal, 1995; Goodhill and Sejnowski, 1996; Kouropteva *et al.*, 2005; Bernatavičienė *et al.*, 2006). It is shown in Karbauskaitė *et al.* (2007) that Spearman’s rho is suitable to estimate the topology preservation after visualizing the data by the LLE algorithm, too. König’s measure is used to estimate the topology preservation of the maps, obtained by self-organizing neural networks (König, 2000; Estevez *et al.*, 2005). In our paper, it is shown that this measure can be also successfully used to estimate the topology preservation after visualizing data by LLE. Mean relative rank errors are used to estimate the topology preservation of the maps, obtained by many nonlinear dimensionality reduction methods for the artificial and real faces (Lee and Verleysen, 2007). In this paper, we have compared these three measures and noticed the advantages of König’s measure and mean relative rank errors compared with Spearman’s rho.

2. Locally Linear Embedding Algorithm (LLE)

It is noted in Yang (2006) that LLE works by assuming that the manifold is well sampled, i.e., there are enough data, and each data point and its neighbours lie on or close to a locally linear patch. Therefore a data point can be approximated as a weighted linear combination of its neighbours. The basic idea of LLE is that such a linear combination is invariant under linear transformations (translation, rotation, and scaling) and, therefore, it should remain unchanged after the manifold has been unfolded to a low space. The low-dimensional configuration of data points is given by solving two constrained least squares optimization problems.

The input of the LLE algorithm consists of m n -dimensional vectors (points) $X_i = (x_{i1}, \dots, x_{in})$, $i = \overline{1, m}$ ($X_i \in R^n$), that are assembled in a matrix \mathbf{X} of size $n \times m$. The output consists of m d -dimensional vectors (points) $Y_i = (y_{i1}, \dots, y_{id})$, $i = \overline{1, m}$ ($Y_i \in R^d$), that are assembled in a matrix \mathbf{Y} of size $d \times m$. The LLE algorithm has three steps. In the first step, we identify k neighbours of each data point X_i . This can be done either by identifying a fixed number of k nearest neighbours of each data point in terms of Euclidean distances or by choosing all points within a ball of a fixed radius. In the second step, we compute the weights w_{ij} that reconstruct each data point X_i best from its neighbours X_{i1}, \dots, X_{ik} , minimizing the following error function:

$$E(W) = \sum_{i=1}^m \left\| X_i - \sum_{j=1}^k w_{ij} X_{ij} \right\|^2, \quad (1)$$

subject to the constraint $\sum_{j=1}^k w_{ij} = 1$. Here $X_{ij} = (x_{i1}^j, \dots, x_{in}^j)$, $i = \overline{1, m}$, $j = \overline{1, k}$ and $\|\cdot\|$ is the Euclidean norm. This is a typical constrained least squares optimization problem that can be easily answered by solving a linear system of equations.

Consider a particular data point X_i with k nearest neighbours X_{ij} and reconstruction weights w_{ij} , $j = \overline{1, k}$ that sum up to one. We can write the reconstruction error as

$$\begin{aligned} E^{(i)}(W) &= \left\| X_i - \sum_{j=1}^k w_{ij} X_{ij} \right\|^2 = \left\| \sum_{j=1}^k w_{ij} (X_i - X_{ij}) \right\|^2 \\ &= \sum_{j,l=1}^k w_{ij} w_{il} c_{jl}^i = \sum_{j=1}^k w_{ij} \sum_{l=1}^k c_{jl}^i w_{il}. \end{aligned} \quad (2)$$

Here $C^i = \{c_{jl}^i\}$, $j, l = \overline{1, k}$ is the $k \times k$ local Gram matrix with the elements defined by the following equation:

$$c_{jl}^i = (X_i - X_{ij}) \cdot (X_i - X_{il}), \quad (3)$$

where X_{ij} and X_{il} are the neighbours of X_i .

LLE may be generalized using other metric distances apart from Euclidean. For example, the kernel distance may be used to find the nearest neighbours in the kernel feature space, instead of finding neighbours in the original input space (as the original LLE

does). Kernel-based learning methods (support vector machines, the kernel PCA and others; Cristianini and Taylor, 2000) are often used in machine learning and data mining. In DeCoste (2001), the use of distances based on Mercer kernels is explored. As a result, a new kernelized form of LLE, called KLLE, has been proposed.

Let $\phi(\cdot)$ be a mapping function from the original n -dimensional space into another high-dimensional, possibly infinite-dimensional feature space. If X_a and X_b are two vectors from R^n , then the kernel is computed as follows:

$$\kappa(X_a, X_b) = \phi(X_a) \cdot \phi(X_b), \quad (4)$$

i.e., $\kappa(X_a, X_b)$ is an inner product of two vectors $\phi(X_a)$ and $\phi(X_b)$ from the kernel feature space without explicitly computing the coordinates of these vectors. In this way, kernels allow large nonlinear feature spaces to be explored avoiding a vast dimensionality. In some cases, explicit computing of the coordinates of vectors in the kernel feature space may be useful (Dzemyda, 2001).

In DeCoste (2001), the various Mercer kernels (the polynomial kernel, the radial basis function kernel (Gaussian kernel), the linear kernel) are applied to LLE. Here the elements of the kernel Gram matrix are defined by the following equation:

$$\begin{aligned} c_{jl}^i &= (\phi(X_i) - \phi(X_{ij})) \cdot (\phi(X_i) - \phi(X_{il})) \\ &= \kappa(X_i, X_i) - \kappa(X_i, X_{ij}) - \kappa(X_i, X_{il}) + \kappa(X_{ij}, X_{il}), \end{aligned} \quad (5)$$

where X_{ij} and X_{il} are the neighbours of X_i .

Just like Roweis and Saul (2000) and Saul and Roweis (2003), we use the linear kernel in this paper: $\phi(X_a) = X_a$ and $\kappa(X_a, X_b) = X_a \cdot X_b$. Application of a nonlinear kernel is the object of future research, based on the results of this paper. It is shown below that even aiming merely at the linear kernel we can achieve good results.

The way to minimize error (2) is simply to solve the linear system of equations

$$\sum_{l=1}^k c_{jl}^i w_{il} = 1, \quad (6)$$

and then to rescale the weights so that they sum up to one:

$$w_{ij} \leftarrow w_{ij} / \sum_{l=1}^k w_{il}. \quad (7)$$

If the Gram matrix is singular or nearly singular – it happens, for example, when there are more neighbours than the analysed data dimensionality ($k > n$), or when the data points are not in the general position – it can be conditioned (before solving the system) by adding a small multiple of the identity matrix

$$c_{jl}^i \leftarrow c_{jl}^i + \delta_{jl} Tr(C^i)t, \quad (8)$$

where $Tr(C^i)$ is the trace of C^i , $\delta_{jl} = 1$ if $j = l$, and 0, otherwise. A control parameter t is set by the user ($t > 0$, $t \ll 1$).

The third step consists in mapping each data point X_i to a low-dimensional point Y_i , which preserves best a high-dimensional neighbourhood geometry represented by the weights w_{ij} . So, the weights are fixed and embedded coordinates Y_i are sought by minimizing the following function:

$$\Phi(Y) = \sum_{i=1}^m \left\| Y_i - \sum_{j=1}^k w_{ij} Y_{ij} \right\|^2 \quad (9)$$

subject to two constraints: $\frac{1}{m} \sum_{i=1}^m Y_i Y_i^T = I$ and $\sum_{i=1}^m Y_i = 0$, where I is the $d \times d$ identity matrix, that provide a unique solution. The most straightforward method for computing the d -dimensional coordinates ($d < n$) is to find the bottom $d + 1$ eigenvectors of the sparse matrix $M = (I - W)^T (I - W)$, where $W = (w_{1j}, w_{2j}, \dots, w_{mj})$, $j = \overline{1, k}$. These eigenvectors are associated with the $d + 1$ smallest eigenvalues of M . The bottom eigenvector, whose eigenvalue is closest to zero, is the unit vector with all equal components and it is discarded. The remaining d eigenvectors form the d embedding coordinates that are found by LLE.

3. Three Topology Preservation Measures

3.1. Spearman's rho

In order to quantitatively estimate the topology preservation, Spearman's rho (Siegel and Castellan, 1988) is commonly used. This quantitative numerical measure estimates the correlation of rank order data, i.e., how well the corresponding low-dimensional projection preserves the order of pairwise distances between the high-dimensional data points converted to ranks. Spearman's rho is computed by using the following equation:

$$\rho_{Sp} = 1 - \frac{6 \sum_{i=1}^T (r_{\mathbf{X}}(i) - r_{\mathbf{Y}}(i))^2}{T^3 - T},$$

where T is the number of distances to be compared ($T = m(m - 1)/2$), $r_{\mathbf{X}}(i)$, $i = \overline{1, T}$ are the ranks (order numbers) of pairwise distances calculated for the original (n -dimensional) data points and sorted in ascending order, and $r_{\mathbf{Y}}(i)$, $i = \overline{1, T}$ are the ranks (order numbers) of pairwise distances calculated for the projected (d -dimensional) data points and sorted in ascending order. As usual, $-1 \leq \rho_{Sp} \leq 1$. The best value of Spearman's rho is equal to one.

It is shown in Karbauskaitė *et al.* (2007) that Spearman's rho is suitable to estimate the topology preservation after visualizing the data by the LLE algorithm. When calculating pairwise distances for the original data points (for $r_{\mathbf{X}}(i)$), it is necessary to use geodesic distances with selected rather a small number of neighbours (≤ 10) for getting values of

these distances (Karbauskaitė *et al.*, 2007). In the case of a projection, the dimensionality of which is the same as the intrinsic dimensionality, both the Euclidean and geodesic distances may be used when calculating pairwise distances for the projected data points (for $r_{\mathbf{Y}}(i)$).

3.2. Konig's Measure (KM)

The topology preservation measure used in Konig (2000) is based on the assessment of rank order in the input and output spaces, too. Let us denote this measure as KM. This measure has two control parameters – numbers of the nearest neighbours: k_1 and k_2 ($k_1 < k_2$). The Euclidean distances estimate a neighbourhood here. Let us denote:

- by $X_{ij}, j = \overline{1, k_1}$, k_1 nearest neighbours of the n -dimensional point X_i , where the distances between X_i and its neighbours satisfy the following inequality $\|X_i, X_{ij_1}\| < \|X_i, X_{ij_2}\|$ with $j_1 < j_2$;
- by $Y_{ij}, j = \overline{1, k_2}$, k_2 nearest neighbours of the d -dimensional point Y_i ;
- by $\bar{r}_{\mathbf{X}}(i, j)$ a rank of the j th neighbour X_{ij} of the point X_i , where the rank means the order number of X_{ij} in the analysed data set $\mathbf{X} = \{X_1, \dots, X_m\}$;
- and by $\bar{r}_{\mathbf{Y}}(i, j)$ a rank of the j th neighbour Y_{ij} of Y_i , corresponding to X_i . Here the rank means the order number of Y_{ij} in the set $\mathbf{Y} = \{Y_1, \dots, Y_m\}$.

The topology preservation measure for the i th point and the j th neighbour is calculated as follows:

$$\text{KM}_{ij} = \begin{cases} 3, & \text{if } \bar{r}_{\mathbf{X}}(i, j) = \bar{r}_{\mathbf{Y}}(i, j), \\ 2, & \text{if } \bar{r}_{\mathbf{X}}(i, j) = \bar{r}_{\mathbf{Y}}(i, l), \quad l = \overline{1, k_1}, \quad j \neq l, \\ 1, & \text{if } \bar{r}_{\mathbf{X}}(i, j) = \bar{r}_{\mathbf{Y}}(i, t), \quad t = k_1 + 1, k_2, \quad k_1 < k_2, \\ 0, & \text{else.} \end{cases}$$

The general measure KM is calculated as follows:

$$\text{KM} = \frac{1}{3k_1 \times m} \sum_{i=1}^m \sum_{j=1}^{k_1} \text{KM}_{ij}.$$

The range of KM is between 0 and 1, where 0 indicates a poor neighbourhood preservation, and 1 indicates a perfect one.

Analysis of the parameters of topology preservation measure KM

The topology preservation measure KM has two control parameters: a smaller number k_1 of the nearest neighbours and a larger number k_2 ($k_1 < k_2$) of the nearest neighbours of each data point. In order to analyse the influence of these parameters on the obtained value of KM, several investigations have been pursued.

The first investigation (Fig. 2) is performed using a nonlinear two-dimensional S-manifold ($m = 1000$) (Fig. 16a). The LLE algorithm has been run for many times gradually

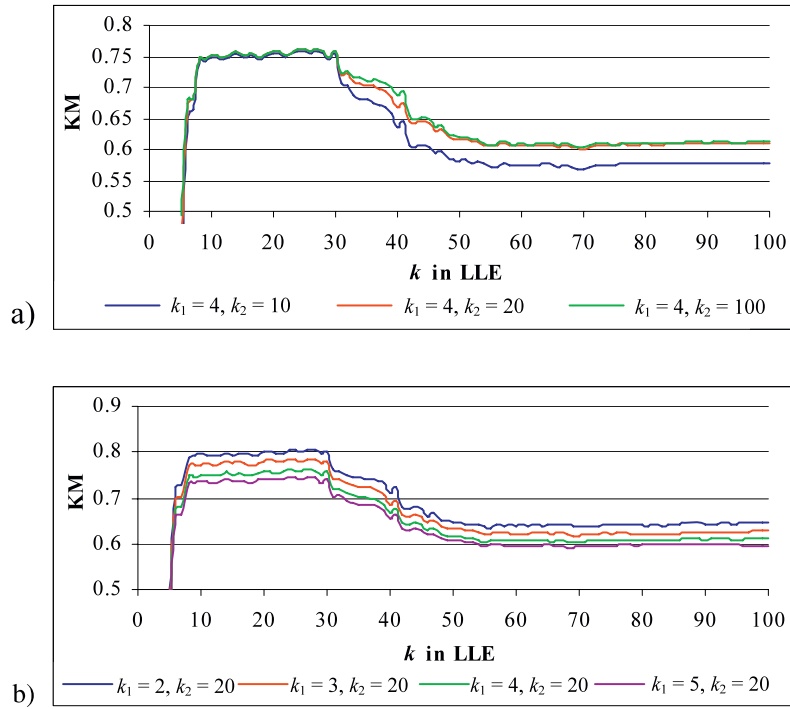


Fig. 2. KM dependences on the LLE parameter k obtained with different combinations of k_1 and k_2 after visualizing the S-manifold by LLE.

increasing the number of neighbours $k \in [5; 100]$ in the LLE algorithm and calculating the topology preservation measure KM for each selected value of k . At first, while calculating the values of KM, a smaller number k_1 of the nearest neighbours was fixed ($k_1 = 4$) and a larger number k_2 of the nearest neighbours was increased little by little ($k_2 = \{10, 20, 100\}$). Three KM dependences on the LLE parameter k have been obtained with different combinations of k_1 and k_2 : $\{4, 10\}$, $\{4, 20\}$ and $\{4, 100\}$. Fig. 2a shows that the values of all dependences are approximately equal, if the obtained value of KM is the best one, i.e., $KM \approx 0.75$. For $k > 30$, the values of KM are decreasing. Besides, if the parameter $k_2 \geq 20$, then the dependences are approximately equal. With $k_2 = 10$, the obtained dependence of KM has lower values than that in dependences with $k_2 = 20$ or $k_2 = 100$. However, the average difference between KM ($k_2 = 10$) and KM ($k_2 = 100$) is only $\approx 6\%$. Thus, the parameter k_2 does not have a great influence while calculating the value of KM.

Subsequently, while calculating the values of KM, the value of k_2 was fixed ($k_2 = 20$) and the value of k_1 was increased little by little; $k_1 = \{2, 3, 4, 5\}$. Four KM dependences on the LLE parameter k have been obtained with different combinations of k_1 and k_2 : $\{2, 20\}$, $\{3, 20\}$, $\{4, 20\}$ and $\{5, 20\}$. We have noticed that the values of KM decrease while increasing the value of k_1 (Fig. 2b). However the parameter k_1 influences only the magnitude of the KM value, while the form of KM dependence on k remains approx-

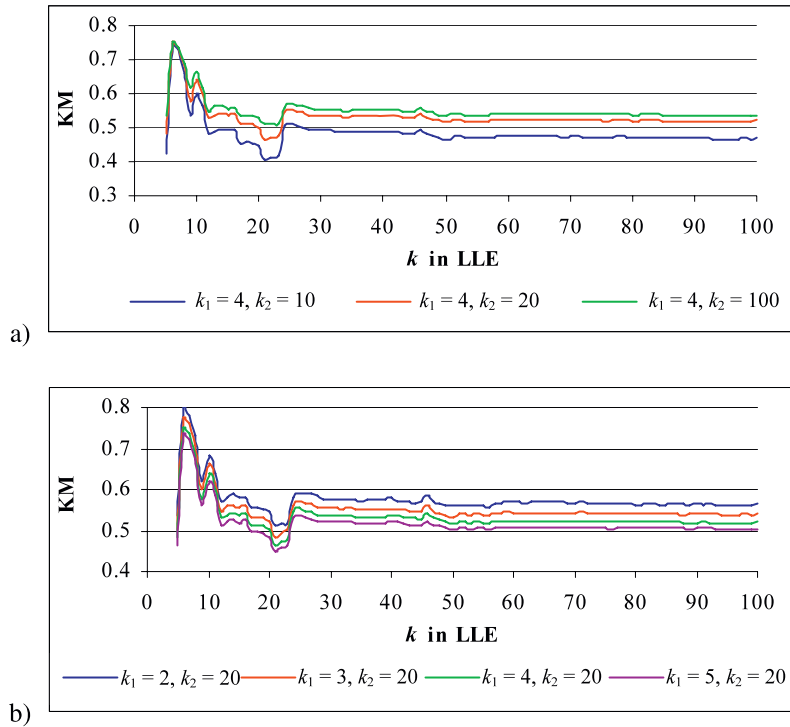


Fig. 3. KM dependences on the LLE parameter k obtained with different combinations of k_1 and k_2 after visualizing the “Swiss roll” by LLE.

imately the same. Therefore, any of these dependences can be used while looking for such a number k (or its interval) of the nearest neighbours in the LLE algorithm that a manifold be successfully unravelled to a lower space.

Analogical investigations have been performed with a nonlinear two-dimensional manifold “Swiss roll” ($m = 1000$) (Fig. 16b). The results obtained are similar to the case of S-manifold (Fig. 3).

3.3. Mean Relative Rank Errors (MRRE)

In Lee and Verleysen (2007), a topology preservation measure, based on the proximity rank, is proposed. The rank $\bar{r}_x(i, j)$ is computed as follows:

- Using the analysed data set \mathbf{X} and taking the i th point X_i as a reference, compute all the Euclidean distances $\|X_i - X_t\|$, for $1 \leq t \leq m, t \neq i$.
- Sort the obtained distances in ascending order, and let the output $\bar{r}_x(i, j)$ be the rank of X_j according to the sorted distances. Note that if $j = \arg \min_{1 \leq t \leq m, t \neq i} \|X_i - X_t\|$, then $\bar{r}_x(i, j) = 1$.

Two different measures, called mean relative rank errors, are calculated as follows:

$$\begin{aligned} \text{a) MRRE}(\mathbf{X} \rightarrow \mathbf{Y}) &= \frac{1}{C} \sum_{i=1}^m \sum_{j \in N_K(X_i)} \frac{|\bar{r}_{\mathbf{X}}(i, j) - \bar{r}_{\mathbf{Y}}(i, j)|}{\bar{r}_{\mathbf{X}}(i, j)}, \\ \text{b) MRRE}(\mathbf{Y} \rightarrow \mathbf{X}) &= \frac{1}{C} \sum_{i=1}^m \sum_{j \in N_K(Y_i)} \frac{|\bar{r}_{\mathbf{X}}(i, j) - \bar{r}_{\mathbf{Y}}(i, j)|}{\bar{r}_{\mathbf{Y}}(i, j)}, \end{aligned}$$

where $N_K(X_i)$ denotes the set of order numbers of K neighbours of X_i . The normalization factor is given by

$$C = m \sum_{l=1}^K \frac{|2l - m - 1|}{l}.$$

It scales the error between 0 and 1. Both measures ($\text{MRRE}(\mathbf{X} \rightarrow \mathbf{Y})$ and $\text{MRRE}(\mathbf{Y} \rightarrow \mathbf{X})$) vanish if the K nearest neighbours of each data point appear in the same order in both spaces. Hence the best value of MRRE is equal to zero.

Analysis of the parameter K of mean relative rank errors

There is only one control parameter, the number K of the nearest neighbours of each data point, in the calculation of mean relative rank errors. With a view to find out the influence of this parameter on the mean relative rank errors ($\text{MRRE}(\mathbf{X} \rightarrow \mathbf{Y})$ and $\text{MRRE}(\mathbf{Y} \rightarrow \mathbf{X})$), an investigation with a nonlinear two-dimensional S-manifold ($m = 1000$) (Fig. 16a) has been performed. The LLE algorithm has been run for many times gradually increasing the number of neighbours $k \in [5; 100]$ and calculating mean relative rank errors for each selected value of k . Figs. 4, 5 show that the values of MRRE increase with an increase in number K of the nearest neighbours of each data point both in n -dimensional (Fig. 4) and d -dimensional (Fig. 5) spaces. However, the form of MRRE dependences on k remains approximately the same. Therefore any of these dependences

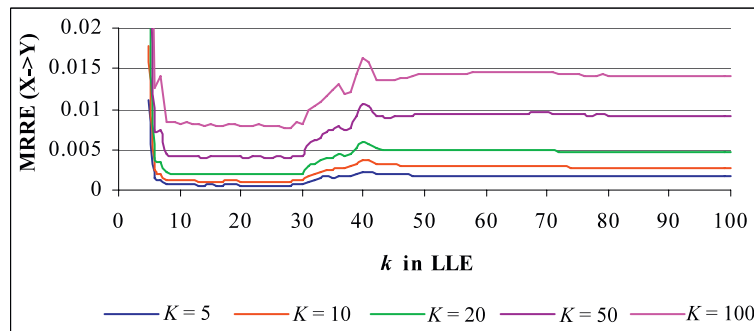


Fig. 4. $\text{MRRE}(\mathbf{X} \rightarrow \mathbf{Y})$ dependences on the LLE parameter k obtained with different values of K in MRRE after visualizing the S-manifold by LLE.

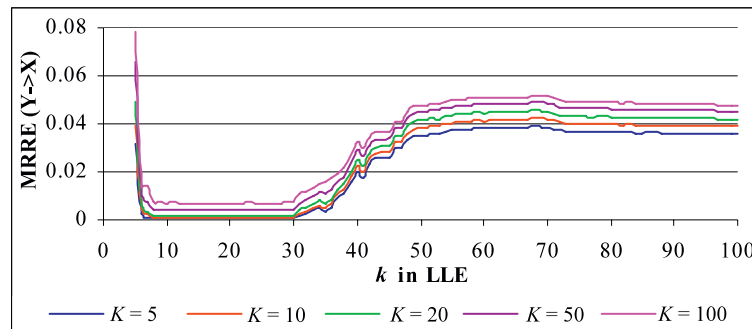


Fig. 5. MRRE($\mathbf{Y} \rightarrow \mathbf{X}$) dependences on the LLE parameter k obtained with different values of K in MRRE after visualizing the S-manifold by LLE.

can be used while looking for such a number k (or its interval) of the nearest neighbours in the LLE algorithm that a manifold be successfully unravelled to a lower space.

4. Comparison of Topology Preservation Measures

In this section, we will compare three topology preservation measures: Spearman’s rho, Konig’s measure (KM), and mean relative rank errors (MRRE). We will also try to find out which measures better estimate the topology preservation of a manifold, after embedding it to a lower d -dimensional space by the LLE algorithm.

To this end, nonlinear 2-dimensional manifolds of various structure and different density are analysed. After getting their projections on a plane by LLE, the values of topology preservation measures are calculated with various values of the LLE parameter k . In this way, dependences of Spearman’s rho, KM and MRRE on k have been obtained.

We state that the chosen value of k in LLE is proper, if LLE succeeds in unravelling the manifold, i.e., LLE perfectly preserves the topology of the manifold.

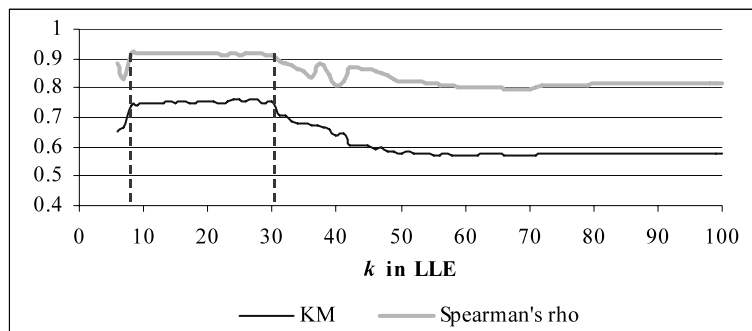


Fig. 6. Dependences of Spearman’s rho and KM on the LLE parameter k obtained after visualizing the S-manifold by LLE.

The first experiment is done using a nonlinear two-dimensional S-manifold ($m=1000$; Fig. 16a). In Karbauskaitė *et al.* (2007), the dependence of Spearman's rho on k has been obtained as well as the manifold successfully unfolded with $k \in [8; 30]$. After calculating the values of KM as $k_1 = 4, k_2 = 10$ for $k \in [6; 100]$, the same proper interval of the nearest neighbours, i.e., $k \in [8; 30]$, was obtained (Fig. 6). We should note the fact that, although the values of Spearman's rho are always higher than KM, it is of great importance that the intervals of the nearest neighbours, corresponding to the best values of both measures, were coincident.

After calculating the values of MRRE ($K = 5$) for $k \in [6; 100]$, we have obtained that the S-manifold was successfully unfolded with $k \in [8; 30]$, too (Fig. 7).

The second investigation is performed with a hemisphere ($m = 294$) (Fig. 16e). Fig. 8 shows that the values of all the three topology preservation measures: Spearman's rho, KM as $k_1 = 4, k_2 = 10$, and MRRE($\mathbf{X} \rightarrow \mathbf{Y}$) as $K = 5$ are the best ones for $k \geq 23$. Thus, the local structure of the hemisphere is unravelled best in this case. Fig. 9 illustrates the visualization of the hemisphere for different values of k .

In practice, nonlinear manifold learning methods are applied, e.g., in image processing. A picture is digitized, i.e., a data point is a vector that consists of colour parameters of pixels, and, therefore, it is of a very large dimension. Often the data are comprised of

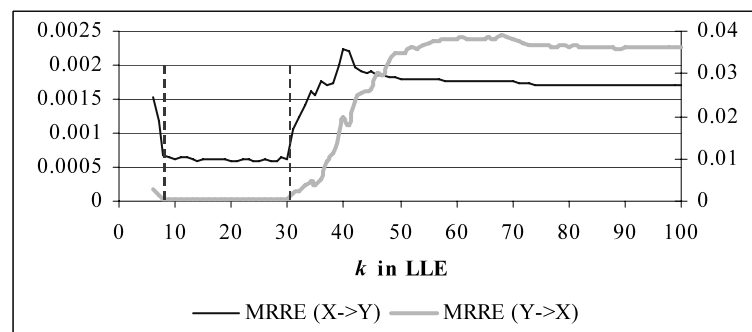


Fig. 7. Dependences of MRRE on the LLE parameter k obtained after visualizing the S-manifold by LLE.

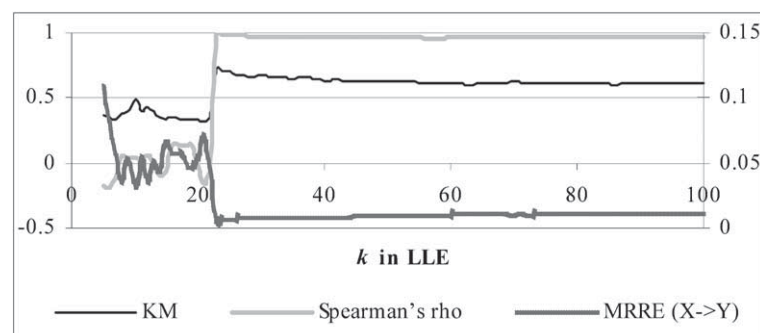


Fig. 8. Dependences of Spearman's rho, KM, and MRRE on the LLE parameter k obtained after visualizing a hemisphere by LLE.

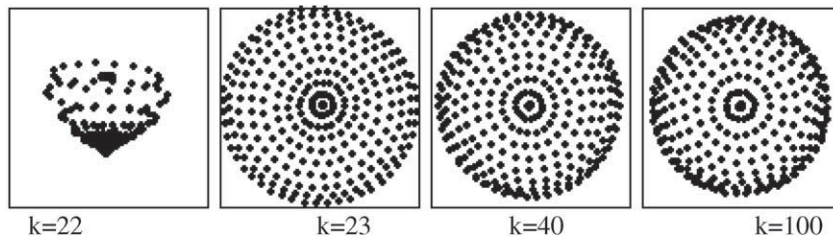


Fig. 9. Projections of a hemisphere on a plane.

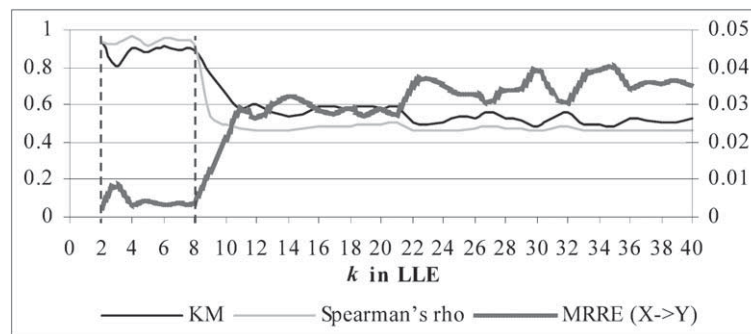


Fig. 10. Dependences of Spearman's rho, KM, and MRRE on the LLE parameter k obtained after visualizing the pictures of a rotating duckling by LLE.

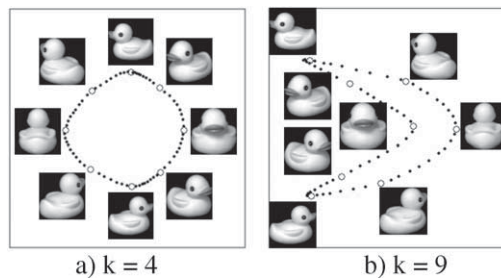


Fig. 11. 2-dimensional embeddings of $m = 72$ pictures of a rotating duckling, obtained by LLE using k nearest neighbours.

pictures of the same object, by turning the object gradually at a certain angle, or taking a picture of the object at different moments, etc. In this way, the points slightly differ from one another, making up a certain manifold. For the third investigation, uncoloured pictures were used, obtained by gradually rotating a duckling at the 360° angle (Nene *et al.*, 1996). The number of pictures was $m = 72$. The images had 128×128 greyscale pixels, therefore the dimensionality of points characterizing each picture in a multidimensional space is $n = 16384$. In Karbauskaitė *et al.* (2007), the dependence of Spearman's rho on k for this data set as well as the right data projections for $k \in [2; 8]$ have been obtained. After calculating the values of KM as $k_1 = 4$, $k_2 = 10$, and $MRRE(\mathbf{X} \rightarrow \mathbf{Y})$

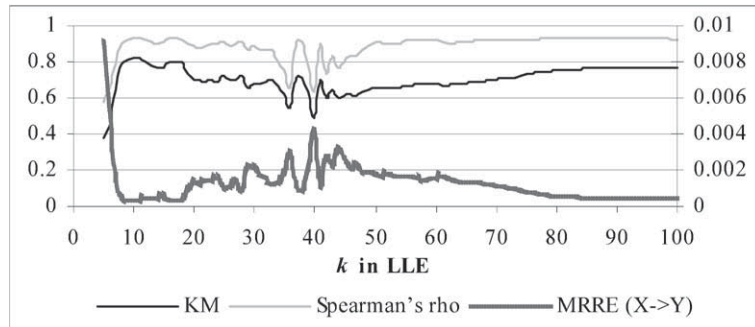


Fig. 12. Dependences of Spearman's rho, KM, and MRRE on the LLE parameter k obtained after visualizing the manifold "Twin peaks" by LLE.

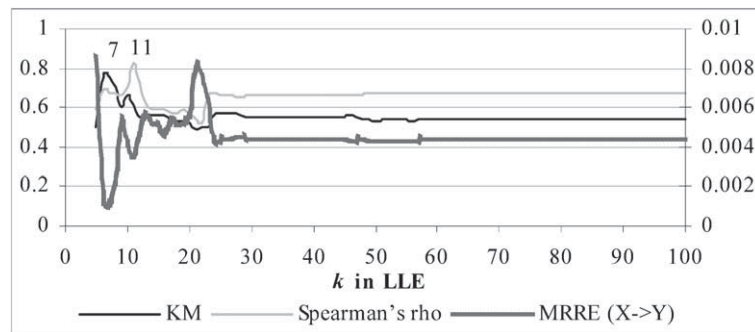


Fig. 13. Dependences of Spearman's rho, KM, and MRRE on the LLE parameter k obtained after visualizing "Swiss roll" by LLE.

as $K = 5$ for $k \in [2; 40]$, the same proper interval of the nearest neighbours as in Spearman's rho case has been obtained (Fig. 10). Fig. 11 illustrates the visualization of pictures of a duckling at different values of k .

Spearman's rho has also been successfully applied to evaluate the topology preservation while investigating a manifold "Twin peaks" ($m = 2000$) (Fig. 16c) in Karbauskaitė *et al.* (2009). We can see from Fig. 12 that, in this case, all the three measures: Spearman's rho, KM as $k_1 = 4$, $k_2 = 20$, and $\text{MRRE}(\mathbf{X} \rightarrow \mathbf{Y})$ as $K = 5$ also acquire their best values approximately in the same intervals of number k of the nearest neighbours.

After investigating these manifolds (S-manifold, hemisphere, pictures of a rotating duckling, "Twin peaks"), we can state that all the three topology preservation measures—Spearman's rho, KM, and MRRE—can be successfully applied to estimate the topology preservation of manifolds, after visualizing them by LLE. However, let us investigate manifolds of a more difficult structure, for example, "Swiss roll", "Punctured sphere" and verify the suitability of the measures in these cases.

The manifold "Swiss roll" ($m = 1000$) (Fig. 16b) is very convoluted. So it is rather difficult to unroll it. Fig. 13 shows that the measures KM as $k_1 = 3$, $k_2 = 20$ and $\text{MRRE}(\mathbf{X} \rightarrow \mathbf{Y})$ as $K = 5$ acquire their best values with $k = 6$, $k = 7$, while the

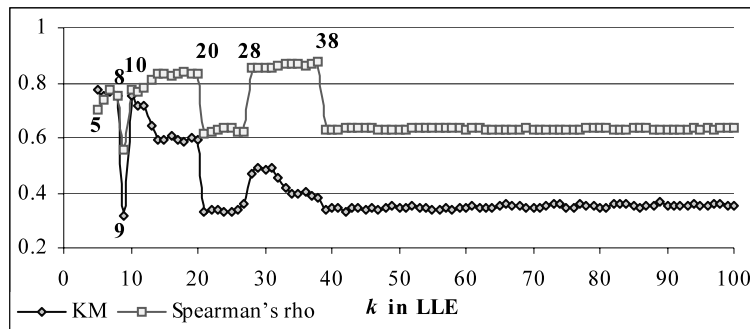


Fig. 14. Dependences of Spearman's rho and KM on the LLE parameter k obtained after visualizing the "Punctured sphere" by LLE.

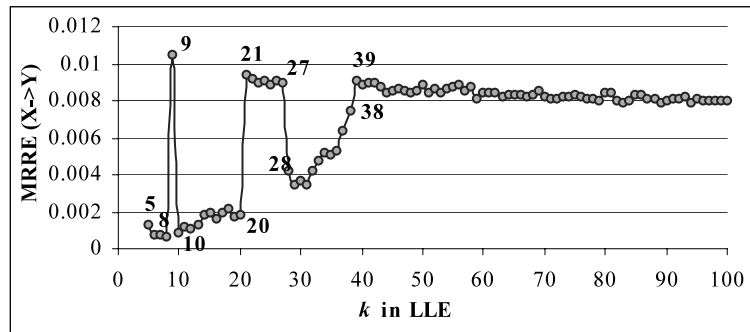


Fig. 15. Dependence of MRRE on the LLE parameter k obtained after visualizing the "Punctured sphere" by LLE.

values of Spearman's rho are not the best ones with these values of the parameter k . However, Spearman's rho acquires the best value if $k = 11$. Meanwhile, KM and MRRE do not have their best values in this case. Such a contradiction means that sometimes conclusions, using different measures, may be different. A question arises: which measure is best? The answer lies in Fig. 17. Obviously, the manifold "Swiss roll" is better unrolled, if $k = 6$, $k = 7$, but not if $k = 11$. Consequently, Spearman's rho is not so good for estimating the topology preservation of this manifold.

The next investigation is pursued with the manifold "Punctured sphere" ($m = 1000$) (Fig. 16d). This manifold is exclusive due to its rather a closed surface (for example, sphere is a fully closed surface), so it is rather complicated to unravel it. We can see from Figs. 14, 15 that, like in the case of "Swiss roll", topology preservation measures yield contrasting results: for $k \in [5; 8] \cup [10; 20] \cup [28; 38]$ the values of KM as $k_1 = 4$, $k_2 = 10$ and $MRRE(\mathbf{X} \rightarrow \mathbf{Y})$ as $K = 5$ gradually become worse, while, on the contrary, the values of Spearman's rho gradually become better, despite that the manifold is unravelled worse and worse. The projections of the "Punctured sphere" are depicted in Fig. 18. Thus, in this case, Spearman's rho is also not so good to assess the topology preservation of the manifold.

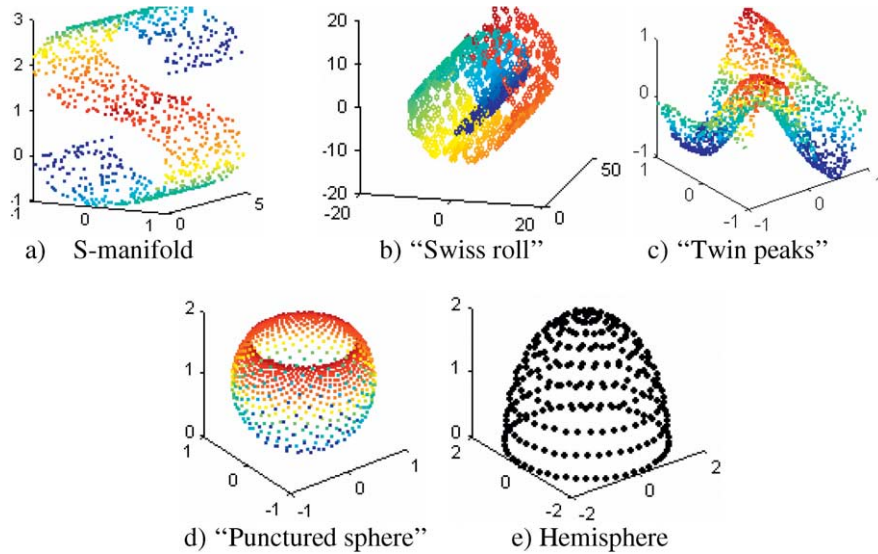


Fig. 16. Nonlinear two-dimensional manifolds.

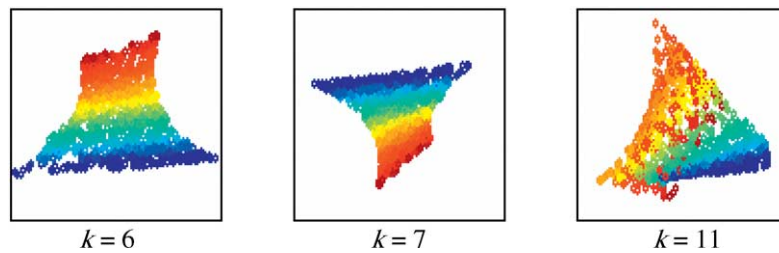


Fig. 17. Projections of a "Swiss roll" on a plane.

One of the main criteria, while estimating algorithms, is the computing time. In order to compare the above-mentioned topology preservation measures in terms of time, the following investigation was performed: various manifolds of different structure and density were analysed, their projections on a plane were obtained by the LLE algorithm and time, necessary to calculate the measures – Spearman's rho, KM, and MRRE – was determined. In Table 1, an average time is given to compute these measures. The values of measures were computed by fixing different number k of neighbours and the used computing time was averaged. It is evident that the KM and MRRE measures are calculated faster than Spearman's rho. Seeking to find a proper number k (or an interval) of the nearest neighbours in LLE, the dependence of the topology preservation measure on different k should be calculated. Usually a wide interval of the nearest neighbours should be analysed, e.g., the values of the parameter k may be taken up to 100. So it is very important to save the computing time as much as possible, while calculating the values of a measure. The shorter calculation time is a great advantage of KM and MRRE as compared with Spearman's rho.

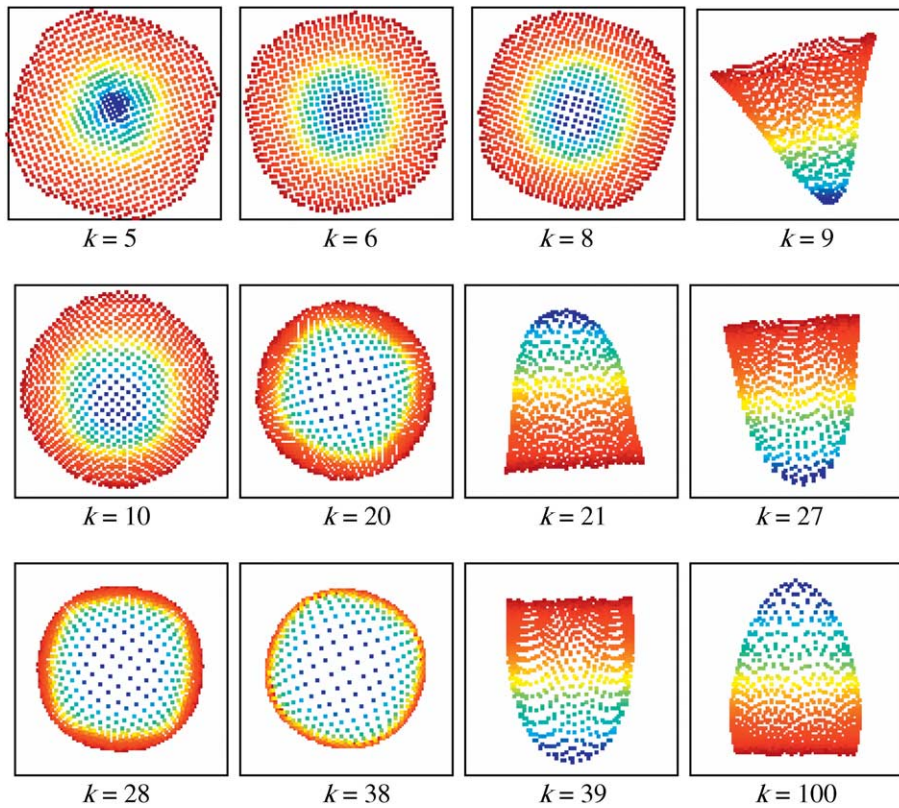


Fig. 18. Projections of a “Punctured sphere” on a plane.

Table 1
Average time to compute the measures

Manifold	Measure		
	KM	MRRE	Spearman’s rho
Pictures of a rotating duckling	0.18 s	0.16 s	0.20 s
Hemisphere	0.05 s	0.04 s	1.36 s
S-manifold	0.46 s	0.44 s	81.61 s
“Swiss roll”	0.48 s	0.45 s	81.81 s
“Punctured sphere”	0.47 s	0.45 s	81.86 s
“Twin peaks”	2.15 s	2.12 s	478.45 s

Conclusions

In data analysis and data visualization, a common goal is to represent data from a high-dimensional space to a low-dimensional space so as to preserve the “internal structure” of

the data in the high-dimensional space as much as possible. Manifold type multidimensional data are often high-dimensional. Research on manifold learning gives a possibility to discover new knowledge about the analysed data set and leads to reasonable decisions. In this paper, we have investigated the LLE (Locally Linear Embedding) method that belongs to a class of nonlinear manifold learning methods that allow unravelling a smooth low-dimensional surface embedded in a higher dimensional space.

In order to quantitatively estimate the topology preservation of a manifold, a quantitative numerical measure must be used. There are lots of different measures of topology preservation. We have investigated three measures: Spearman's rho, Konig's measure (KM), and mean relative rank errors (MRRE). Two criteria were used in the comparative analysis of the three topology preservation measures – the topology preservation quality and computational expenditure.

After investigating the manifolds of a simpler structure, such as an S-manifold, a hemisphere, a manifold "Twin peaks", real pictures, we have noticed that all the three measures – Spearman's rho, KM, and MRRE – can be successfully applied to estimate the topology preservation of manifolds, after visualizing the manifolds by LLE. However, after investigating the manifolds of more complex structures, such as "Swiss roll", "Punctured sphere", we have determined that only KM and MRRE are better for estimating the topology preservation of these manifolds.

Calculation of KM and MRRE is faster because these criteria use the Euclidean distances only, while Spearman's rho uses the geodesic distances that are more computationally expensive. Moreover, KM and MRRE evaluate a limited number of neighbours of each point from the analysed data set. Spearman's rho considers distances between all the pairs of points from the analysed data set. Therefore, it tries to take into account the global structure of the manifold. However, in some cases, it may be not optimal, because some local properties of the manifold may be lost.

Acknowledgements

The authors are very grateful to assoc. prof. dr. Olga Kurasova, Institute of Mathematics and Informatics, for her valuable remarks that allowed us to improve the quality of this paper. The research is partially supported by the Lithuanian State Science and Studies Foundation project "Information technology tools of clinical decision support and citizens wellness for e.Health system (No. B-07019)".

References

- Belkin, M., and P. Niyogi (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, **15**(6), 1373–1396.
- Bernatavičienė, J., G. Dzemyda, O. Kurasova and V. Marcinkevičius (2006). Optimal decisions in combining the SOM with nonlinear projection methods. *European Journal of Operation Research*, **173**, 729–745.
- Bezdek, J.C., and N.R. Pal (1995). An index of topological preservation for feature extraction. *Pattern Recognition*, **28**(3), 381–391.

- Cristianini, N., and J.S. Taylor (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge.
- DeCoste, D. (2001). Visualizing Mercer kernel feature spaces via kernelized locally-linear embeddings. In *8th International Conference on Neural Information Processing (ICONIP 2001)*. Shanghai, China.
- Donoho, D.L., and C. Grimes (2005). Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, **102**(21), 7426–7431.
- Dzemyda, G. (2001). Visualization of a set of parameters characterized by their correlation matrix. *Computational Statistics and Data Analysis*, **36**(10), 15–30.
- Estevez, P.A., C.J. Figueroa and K. Saito (2005). Cross-entropy embedding of high-dimensional data using the neural gas model. *Neural Networks*, **18**(5–6), 727–737.
- Goodhill, G.J., and T.J. Sejnowski (1996). Quantifying neighbourhood preservation in topographic mappings. In *Proceedings of the Third Joint Symposium on Neural Computation*, Vol. 6. California Institute of Technology, Pasadena. pp. 61–82.
- Jolliffe, I.T. (1989). *Principal Component Analysis*. Springer-Verlag, New York.
- Karbauskaitė, R., O. Kurasova and G. Dzemyda (2007). Selection of the number of neighbours of each data point for the locally linear embedding algorithm. *Information Technology and Control*, **36**(4), 359–364.
- Karbauskaitė, R., G. Dzemyda and V. Marcinkevičius (2009). Dependence of locally linear embedding on the regularization parameter. *TOP, An Official Journal of the Spanish Society of Statistics and Operations Research* (accepted).
- Konig, A. (2000). Interactive visualization and analysis of hierarchical neural projections for data mining. *IEEE Transactions on Neural Networks*, **11**(3), 615–624.
- Kouropteva, O., O. Okun and M. Pietikainen (2005). Incremental locally linear embedding algorithm. In *Proceedings of Fourteenth Scandinavian Conference on Image Analysis, Lecture Notes in Computer Science*, Vol. 3540. Springer, Berlin. pp. 521–530.
- Lee, J.A., and M. Verleysen (2007). *Nonlinear Dimensionality Reduction*. Springer, New York.
- Levina, E., and P.J. Bickel (2005). Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, Vol. 17. MIT Press, Cambridge.
- Nene, S.A., S.K. Nayar and H. Murase (1996). *Columbia Object Image Library (COIL-20)*. Technical Report CUCS-005-96.
- Roweis, S.T., and L.K. Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323–2326.
- Saul, L.K., and S.T. Roweis (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J. Machine Learning Research*, **4**, 119–155.
- Siegel, S., and N.J. Castellan (1988). *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. New York, McGraw-Hill.
- Tenenbaum, J.B., V. Silva and J.C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319–2323.
- Venna, J., and S. Kaski (2001). Neighborhood preservation in nonlinear projection methods: An experimental study. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN), Lecture Notes in Computer Science*, Vol. 2130. Springer, Berlin. pp. 485–491.
- Yang, L. (2006). Building connected neighborhood graphs for locally linear embedding. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Vol. 4. IEEE Computer Society, Washington. pp. 194–197.
- Yin, J., D. Hu and Z. Zhou (2007). Growing locally linear embedding for manifold learning. *Journal of Pattern Recognition Research*, **2**(1), 1–16.
- Zhang, Z., and H. Zha (2004). Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment. *SIAM Journal of Scientific Computing*, **26**(1), 313–338.

R. Karbauskaitė doctoral student, engineer programmer, Department of System Analysis, Institute of Informatics and Mathematics. Junior lecturer, Department of Informatics, Vilnius Pedagogical University. Bachelor's degree in Mathematics and Informatics (2003), Master's degree in Informatics (2005) from Vilnius Pedagogical University. Research interests: visualization of multidimensional data, estimation of the visualization quality, data clustering.

G. Dzemyda graduated from Kaunas University of Technology, Lithuania, in 1980, and in 1984 received there the doctoral degree in technical sciences (PhD) after post-graduate studies at the Institute of Mathematics and Informatics, Vilnius, Lithuania. In 1997 he received the degree of Doctor Habilius from Kaunas University of Technology. He was conferred the title of a professor (1998) at Kaunas University of Technology. He is the director of the Institute of Mathematics and Informatics and heads the System Analysis Department of the institute. The areas of research are the theory, development and application of optimization, and the interaction of optimization and data analysis. The interests include optimization theory and applications, data mining in databases, multiple criteria decision support, neural networks, parallel optimization, internet databases, and data visualization.

Topologijos išlaikymo matai daugdaros tipo daugiamatųjų duomenų vizualizavime

Rasa KARBAUSKAITĖ, Gintautas DZEMYDA

Dažnai praktiniuose uždaviniuose sukaupiamų daugiamatųjų duomenų taškai yra išsidėstę ant mažesnio matmenų skaičiaus daugdaros, įterptos į labai didelio skaičiaus matmenų erdvę. Netiesiniai daugdaros atpažinimo metodai automatiškai atranda mažesnio matmenų skaičiaus netiesinę daugdarą didelio skaičiaus matmenų erdvėje ir tada transformuoja tuos duomenų taškus į mažesnio skaičiaus matmenų erdvę, išsaugodami lokalią duomenų struktūrą. Vienu tokių metodų – lokaliai tiesinio vaizdavimo metodu – remiasi šio straipsnio tyrimai. Norint kiekybiškai įvertinti daugdaros topologijos išlaikymą, „išvyniojus“ ją mažesnio skaičiaus matmenų erdvėje, reikia naudoti kiekybinius skaitinius matavimus. Topologijos išlaikymui įvertinti sukurta daugybė įvairių matavimų. Straipsnyje tiriami trys matai. Pirmasis vertina visus atstumus tarp tiriamos duomenų aibės taškų, o kiti du – tarp nedidelio kaimyninių taškų skaičiaus. Parodyta, kad antru atveju matai visada gerai nusako daugdaros topologijos išlaikymą po transformacijos.