Liang Meng
Piotr Breitkopf

INTRINSIC DIMENSIONALITY DETECTION CRITERION BASED ON LOCALLY LINEAR EMBEDDING

Abstract *In this work, we revisit the Locally Linear Embedding (LLE) algorithm that is widely employed in dimensionality reduction. With a particular interest to the correspondences of the nearest neighbors in the original and embedded spaces, we observe that, when prescribing low-dimensional embedding spaces, LLE remains merely a* weight-preserving *rather than a neighborhood-preserving algorithm. Thus, we propose a "neighborhood-preserving ratio" criterion to estimate the minimal intrinsic dimensionality required for neighborhood preservation. We validate its efficiency on sets of synthetic data, including S-curve, Swiss roll, and a dataset of grayscale images.*

Keywords LLE, dimensionality reduction, intrinsic dimensionality, neighborhood preserving

Citation Computer Science 19(3) 2018: 345–356

1. Introduction

Handling high-dimensional data is inconvenient due to the exponential increase of its computational complexity and prohibitive storage. However, real-life data does not necessarily fill high-dimensional space $\mathbb{R}^D$ uniformly. Instead, it is observed to concentrate on underlying manifold $\mathcal{M}$ of a much lower dimensionality, $\mathcal{M} \in \mathbb{R}^d$, $d \ll D$. Thus, dimensionality reduction is one of the challenges usually encountered in statistical pattern recognition, information processing, and scientific computing.

For decades, a large number of algorithms were proposed to seek the most compact embeddings of original data in lower-dimensional spaces. Locally Linear Embedding [5], among others, is an unsupervised eigenvector method that discovers the underlying non-linear structures of the original data. Unlike Metric Multidimensional Scaling (MDS) [2] and Principal Component Analysis (PCA) [3] (which are both linear algorithms), LLE is widely employed, owing not only to its simplicity of implementation but also its capacity of generating highly nonlinear embeddings.

Despite its popularity in dimensionality reduction, LLE does not provide an estimation of the *intrinsic dimensionality*, d_{in}; i.e., the minimal value of d required to represent the data in the target feature space without information loss. The usual criteria based on eigenvalues of cost matrix [4] were found unreliable and worked only for contrived examples where the data lies either essentially on a linear manifold or is sampled basically in a uniform way [6]. In most cases, the spectrum of the cost matrix does not feature a "telltale gap" allowing us to detect the intrinsic dimension. Therefore, d has to be provided by the user as an input parameter to the LLE algorithm [7].

In this work, we observe that the LLE algorithm relies on the assumption that the k nearest neighbors in the original space remain closest in the feature space. However, as LLE is based solely on preserving weights, this assumption does not necessarily hold. With this in mind, we search for a drastic change in the set of closest neighbors when imposed dimensionality d increases rather than looking for a "telltale gap" in the spectrum of the eigenvalues. From a quantitative point of view, we define the neighborhood-preserving ratio as the percentage of the original nearest neighbors within the k nearest neighbors in the embedded space. Finally, we propose this criterion to infer the intrinsic dimensionality.

The remainder of the paper is organized in the following manner: the LLE method is briefly recalled in Section 2, followed by observations on the algorithm and the proposition of the intrinsic dimensionality estimation criterion in Section 3. In Section 4, several benchmark cases with synthetic data validate the proposed criterion. Finally, we close the paper with concluding comments and suggestions for future work in Section 5.

2. Locally linear embedding

Like many other algorithms, LLE discovers the non-linear structure of high-dimensional data by exploiting the local symmetries of linear reconstruction. In this section, we shall briefly recall the principle idea of LLE. It is noticed that, unless otherwise specified, the original high-dimensional data is denoted by matrix $\boldsymbol{X}$, and the embedded $\boldsymbol{Y}$. Two common steps covered in LLE include the following:

1. learn the local geometry around each point,
2. embed high-dimensional data into low-dimensional feature space using local information learned from 1.

In the first step, the intrinsic geometric properties of high-dimensional data are characterized by weight matrix $\boldsymbol{W}$, where each row $\boldsymbol{W}^{(i)}$ minimizes the reconstruction error of point $\boldsymbol{X}^{(i)}$ by its neighbors.

$$\boldsymbol{W}^{(i)} = \underset{\omega_{ij}}{\text{Argmin}} |\boldsymbol{X}^{(i)} - \sum_j \omega_{ij} \boldsymbol{X}^{(j)}|^2, \tag{1}$$

where $\omega_{ij} = 0$ if point $\boldsymbol{X}^{(j)}$ is not one of the k nearest neighbors of point $\boldsymbol{X}^{(i)}$. Note that this weigh matrix is not necessarily symmetrical for two reasons: on the one hand, if $\boldsymbol{X}^{(i)}$ is one of the nearest neighbors of $\boldsymbol{X}^{(j)}$, this does not guarantee that $\boldsymbol{X}^{(j)}$ also lies in the k-neighborhood of $\boldsymbol{X}^{(i)}$, and on the other hand, even $\boldsymbol{X}^{(i)}$ and $\boldsymbol{X}^{(j)}$ are mutually nearest neighbors, the corresponding weights (ω_{ij} and ω_{ji}) probably do not have the same value. By carefully choosing the number of nearest neighbors k, $\boldsymbol{W}$ shall interpret the local properties of the original high-dimensional data, and by design, embedded low-dimensional data $\boldsymbol{Y}$ is interpolated by the same weight matrix $\boldsymbol{W}$, thus leading to the following:

$$\boldsymbol{Y} = \underset{\boldsymbol{Y}^{(i)}}{\text{Argmin}} |\boldsymbol{Y}^{(i)} - \sum_j \omega_{ij} \boldsymbol{Y}^{(j)}|^2. \tag{2}$$

The cost function in the above minimization problem essentially defines a quadratic form.

$$\varepsilon(\boldsymbol{Y}) = \sum_{ij} \boldsymbol{M}_{ij} (\boldsymbol{Y}^{(i)} \cdot \boldsymbol{Y}^{(j)}), \tag{3}$$

in which the components of $\boldsymbol{M}$ are

$$\boldsymbol{M}_{ij} = \delta_{ij} - \omega_{ij} - \omega_{ji} + \sum_l \omega_{li}\omega_{lj}, \tag{4}$$

where $\delta_{ij} = 1$ if $i = j$; otherwise, $\delta_{ij} = 0$. By solving Eq.(2), LLE then finds the lower-dimensional embedding system by computing the bottom $d+1$ eigenvectors of $\boldsymbol{M}$, where d is the desired/prescribed dimensionality of the embedded space. Discarding the bottom eigenvector (the corresponding eigenvalue is zero), the remaining d non-zero eigenvectors provide an ordered set of orthogonal coordinates of the original data in the feature space.

3. Estimation of dimensionality based on LLE

3.1. Role of imposed dimensionality d

In the original version of LLE proposed by Saul and Roweis (2000), the two independent parameters involved in LLE are the number of nearest neighbors per data point k and embedding space dimensionality d.

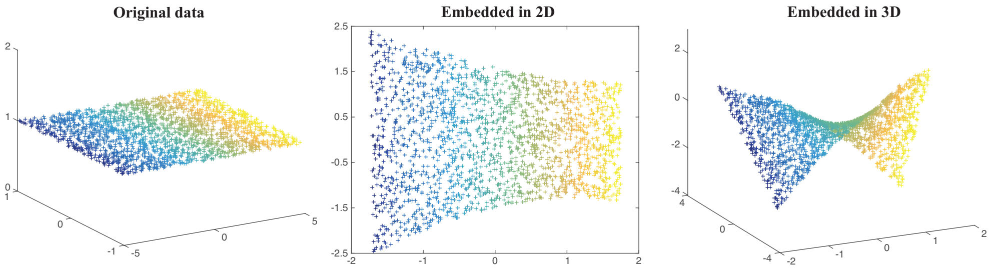

Figure 1. (Left) Original 3D data distributed along 2D manifold.
(Middle) Embedded coordinates of original points using LLE with $d = 2$ and $k = 12$.
(Right) Embedded coordinates of original points using LLE with overestimated embedded dimensionality $d = 3$

We underline that a manifold's intrinsic dimensionality d_{in} may be known a priori in some applications, while in the great majority of other cases, d_{in} is unknown and the users wish to bias the embedding of a particular dimensionality. Unfortunately, the intrinsic dimensionality can not be readily estimated in reality, and an underestimate of embedded dimensionality ($d < d_{in}$) will easily screw up the results since it is impossible to properly describe data of a high dimensionality in a low-dimensional space. Moreover, even for the case where d is overestimated, i.e., $d > d_{in}$, it was reported that LLE might behave pathologically [4]. Figure 1 is such an example where extraneous information was "added" to the intrinsically 2D linear manifold, making it non-linear.

In summary, an improperly chosen prescribed dimensionality can lead to the underperformance of LLE, and it is of vital importance to know the intrinsic dimensionality d_{in} of the original data before employing the algorithm.

3.2. Weight preserving

Another issue concerning LLE is its neighborhood-preserving property. Saul and Roweis (2000) considered LLE as a neighborhood-preserving approach; that is to say, the original data points within a typical vicinity probably remain in the same nearest neighborhood in the embedded system. Despite being the essential intuition of LLE, this neighborhood-preserving property was neither guaranteed by Eq.(1) nor Eq.(2).

Figure 2 shows the embedded coordinates of the benchmark S-curve by employing LLE. The neighbors of an arbitrarily chosen point (marked by the red asterisk) were compared before and after embedding; these are indicated by blue and red circles, respectively. It is found that, even for such a simple 3D example, the neighborhood connectivity of a given point is not 100% guaranteed, and for all of the $k = 12$ neighbors, only six of them were preserved by the embedding (black diamond).

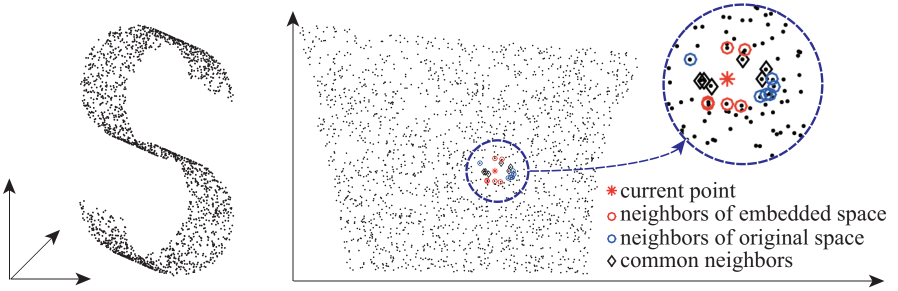

Figure 2. Alteration of neighbors of arbitrary point after embedding using LLE

Here, one should be aware of the difference between *weights* and *neighborhoods*. During the first step of the implementation of LLE, the weights are calculated by reconstructing a given point in the original space with its nearest k neighbors (see Eq.(1)) in the aim of characterizing the local properties of the original data. However, in the embedded space, these weights are linear coefficients corresponding to a group of embedded points, which belong to the same neighborhood in the original instead of the embedded space. In this manner, the neighborhood is not guaranteed in the lower-dimensional space. Thus, we consider the basic LLE as *a weight-preserving rather than a neighborhood-preserving algorithm.*

3.3. Evaluation of intrinsic dimensionality

In view of the weight-preserving rather than the neighborhood-preserving property of LLE, we expect to have some knowledge on the validity/faithfulness of the embedding by quantifying the alterations/evolution of the neighborhoods of all data points. Two coordinate systems (X and Y) are thus defined to refer to the initial and embedded system, respectively.

Suppose $\boldsymbol{V}^{(i)}$ collects all of the reference numbers of the neighbors of $\boldsymbol{X}^{(i)}$. We first define operation $diff(\cdot)$ between $\boldsymbol{V}^{(i)}$ and $\boldsymbol{V}^{(j)}$, which returns the number of common components in the two vectors. As a consequence, $diff(\boldsymbol{V}_{\mathrm{X}}^{(i)}, \boldsymbol{V}_{\mathrm{Y}}^{(i)})$ provides the number of preserved neighbors for point $\boldsymbol{X}^{(i)}$ in two systems. For example, $diff(\boldsymbol{V}_{\mathrm{X}}^{(1)}, \boldsymbol{V}_{\mathrm{Y}}^{(1)}) = diff(\boldsymbol{V}_{\mathrm{X}}^{(2)}, \boldsymbol{V}_{\mathrm{Y}}^{(2)}) = 2$ as illustrated in Figure 3. Similarly, this operation could be extended to two matrices.

$$diff(\boldsymbol{V}_{\mathrm{X}}, \boldsymbol{V}_{\mathrm{Y}}) = \sum_{l=1}^{N} diff(\boldsymbol{V}_{\mathrm{X}}^{(l)}, \boldsymbol{V}_{\mathrm{Y}}^{(l)}) \tag{5}$$

where N is the number of sample points at hand or the column number of both matrices.

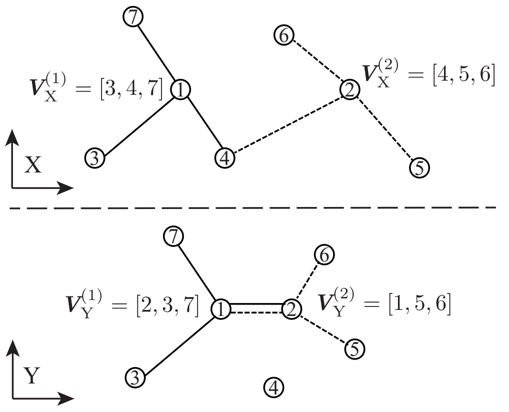

Figure 3. Connectivity/vicinity of points before and after embedding

Based on Eq.(5), we consequently propose the concept of a "neighborhood-preserving ratio" that quantifies the evolution of a neighborhood before and after embedding.

$$\gamma(d) = \frac{diff(\boldsymbol{V}_{\mathrm{X}}, \boldsymbol{V}_{\mathrm{Y}})}{k \cdot N} \times 100\%. \tag{6}$$

Quite obviously, the value of γ varies between 0% and 100%; in general, the larger this value is, the better preserved the neighborhood is. It is logical (and certainly intuitive) to expect that the neighborhoods of each point change slightly if the embedding persists the original information, while they change dramatically if the embedded dimensionality is wrongly chosen. By "wrongly," we mean that embedded dimensionality d was too small to describe the high-dimensional data properly or, to be simple, $d < d_{in}$. With this consideration, a "telltale gap" may be observed between $\gamma(d)$ and $\gamma(d+1)$, indicating that the original data lies in a $d-$dimensional space. We consequently summarize that the "neighborhood-preserving ratio" provides us with an estimate of the intrinsic dimensionality of the original data.

Another way of evaluating the intrinsic dimensionality using the proposed criterion may consist of choosing a proper threshold value ε for γ. By increasing the prescribed dimensionality, the intrinsic dimensionality of the original data points is estimated by d until $\gamma(d) > \varepsilon$. However, the choice of the threshold value may be problem-dependent; an empirical value will be found in the following section, referring to several benchmark cases.

4. Examples

4.1. Examples on artificial noiseless data

We choose several manifolds embedded initially in 3D space, including an S-shape curve, a Swiss roll, and a flat surface for verifying the efficiency of the proposed criterion. The first two manifolds are benchmarks of nonlinear dimensionality reduction, while the third one is proposed for comparison purposes. For all three cases, the input of LLE consists of $N = 2000$ points sampled randomly from the continuous 3D surfaces in Figure 4.

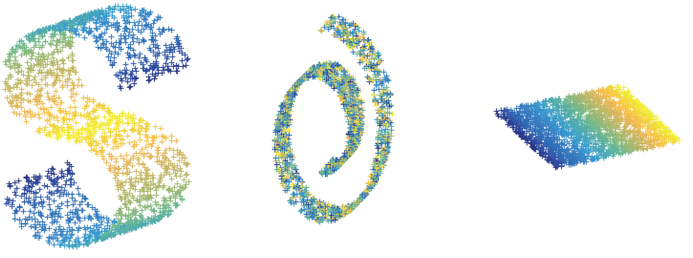

Figure 4. Original data sampled from 3D manifolds: S-shape curve (left); Swiss roll (middle); flat surface (right)

By following Eqs.(1)–(4), LLE is performed blindly with imposed dimensionality d varying form 1 to 10 without knowing a priori the intrinsic dimensionality. Figures 5–7 present the corresponding neighborhood-preserving ratios γ with respect to different embedded dimensionalities d.

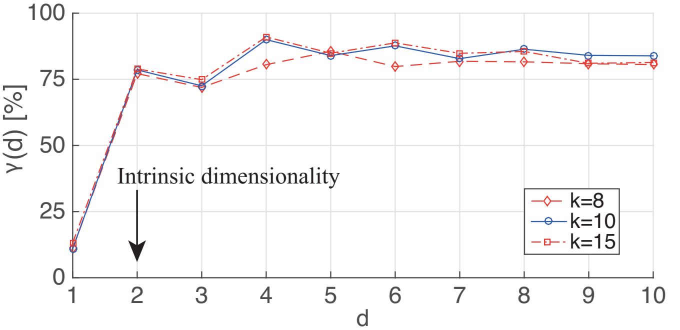

Figure 5. Neighborhood-preserving ratios for S-shape curve

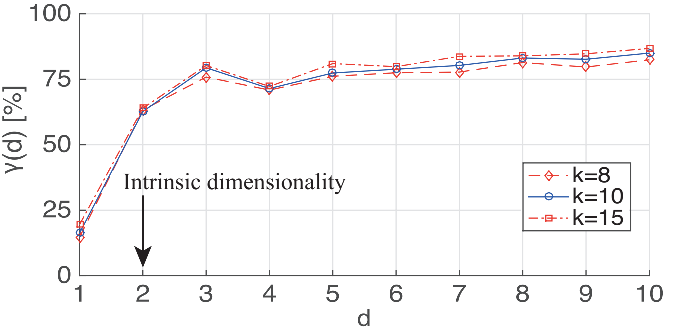

Figure 6. Neighborhood-preserving ratios for Swiss roll

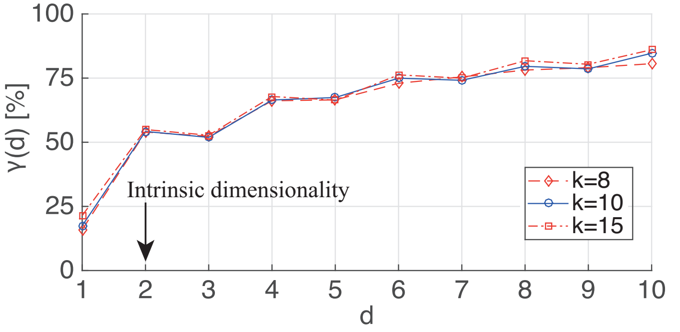

Figure 7. Neighborhood-preserving ratios for plane surface

As clearly indicated in the three $d - \gamma(d)$ diagrams, a telltale gap can always be observed between $\gamma(1)$ and $\gamma(2)$ regardless of the number of neighbors chosen ($k = 8, 10, 15$). This should demonstrate that the intrinsic dimensionality of each manifold is $d_{in} = 2$.

One may argue that embedding this 3D data in higher-dimensional spaces ($d > 3$) makes little sense since the dimensionality of the original data points is 3 and we are certain that d_{in} shall never exceed this value. In this work, the ratios of the neighborhood preserving corresponding to the higher-dimensional embedded spaces were provided with the aim of showing that the neighborhood-preserving ratios stabilized when the embedded dimensionality exceeds the intrinsic one.

Moreover, we noticed in some cases that we are probably yielding an even worse embedding while increasing the embedded dimensionality; e.g., $\gamma(3) < \gamma(2)$ as observed in Figure 5 and Figure 7. This observation in fact does not impose any restriction on the estimation of intrinsic dimensionality, and it confirms the numerical instability of LLE stated in Section 3.1.

Finally, we list the neighborhood-preserving ratios for $d = 1, 2, 3$ in Table 1 while setting $k = 12$; the aim of this is to find an empirical threshold for the criterion. Despite the three manifolds shown in Figure 4, two other supplemental examples (i.e., a partial sphere [an intrinsic 2D manifold] and a straight line [an intrinsic 1D manifold]) are provided.

Table 1
Intrinsic dimensionality detected by LLE-based criterion ($k = 12$)

	$d = 3$	$d = 2$	$d = 1$	d_{in}	reference value
S-curve	74.85%	79.48%	12.84%	2	2
Swiss roll	77.58%	65.92%	18.50%	2	2
Plane surface	52.65%	59.40%	18.99%	2	2
Partial Sphere	90.22%	85.85%	15.85%	2	2
Straight line	99.81%	99.55%	99.97%	1	1

The readers are invited to refer to [6] for more details about these manifolds. Except for the coincidence of the estimated and reference intrinsic dimensionality, we conclude from Table 1 that 50% may be a reasonable value for criterion γ. (Informally, we recommend the embedding as "faithful" if the majority of the topological relationships remain unchanged.) Therefore, intrinsic dimensionality d of the original data can also be evaluated if $\gamma(d) > 50\%$ *and* $\gamma(d-1) < 50\%$. However, we will later find this threshold value problem-dependent, and it would be more practical to combine it with the telltale gap observed in the $d - \gamma(d)$ diagram.

4.2. Robustness verification on noisy data

This section focuses on the stability of the criterion when applied to noisy data. The S-shape was adopted for illustration purposes, and a zero-mean normally distributed noise was added to the coordinates of the points. The standard deviation of the noise was set to 2% of the smallest dimension of the bounding box that encloses the entire data set.

As in Figure 8, the intrinsic dimensionality of the noisy data was estimated to be $d_{in} = 3$ when $k = 8$ neighbors were considered while $d_{in} = 2$ for the other two cases. Clearly, there was an overestimation when fewer neighbors were considered. Fortunately, this phenomenon could likely be alleviated by including more neighbors.

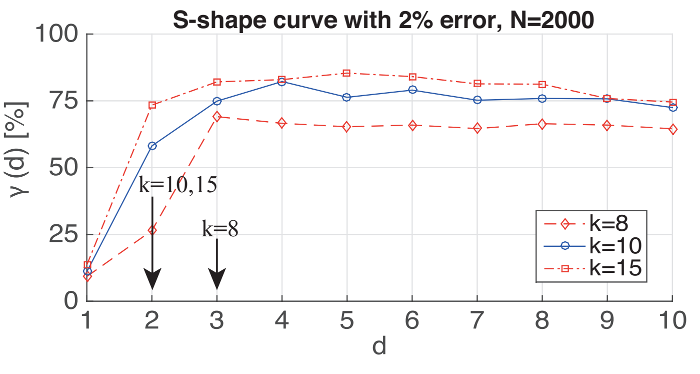

Figure 8. Neighborhood-preserving ratios for noisy S-shape curve

On the other hand, it is proposed in [1] that a larger number of data points or a lesser-curved manifold can render a substantially better degree of noise tolerance. Despite being reported with regard to ISOMAP (another non-linear dimensionality reduction algorithm), this phenomenon will be validated for LLE. To this end, we sampled the same manifold again but more densely (i.e., $N = 4000$) with the same level of 2% error. The $d - \gamma(d)$ diagram is presented in Figure 9. It can be noticed that the intrinsic dimensionality is accurately estimated even with $k = 8$, and the proposed criterion is quite stable for noisy data.

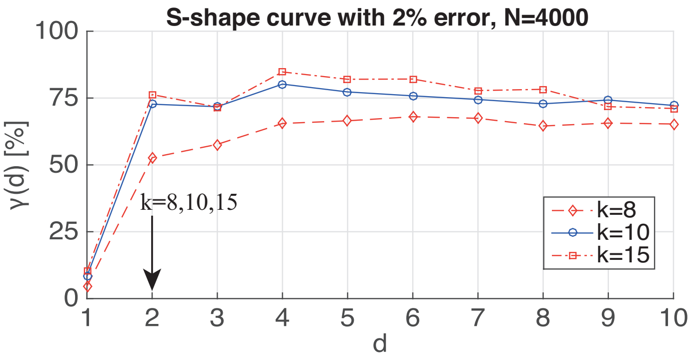

Figure 9. Neighborhood-preserving ratio for noisy S-shape curve

4.3. Example on grayscale images

By now, the proposed criterion has been validated on both noiseless and noisy data and was proven to be stable. However, the original data involved in these examples is always three-dimensional, and none of it is initially expanded in high-dimensional space. In this part, a series of grayscale images are chosen as input data to verify the validity of the criterion on high-dimensional data.

The dataset consists of 500 grayscale images of a black square moving in a more extensive domain. With 113 pixel points along each direction, the original dimensionality of the image was $D = 12769$. It is known a priori that all of these images should lie on a constrained manifold, parameterized by two variables (x_0, y_0) defining its position plus rotation angle θ, Figure 10; thus, the underlying intrinsic dimensionality of these images should be $d_{in} = 3$.

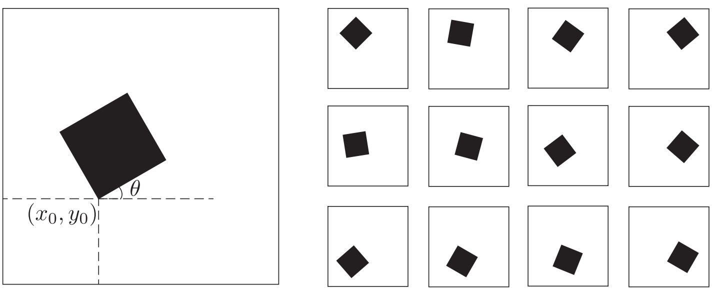

Figure 10. Grayscale images defined on three parameters

In Figure 11, we figured out that 50% may become unreliable as the threshold value for intrinsic dimensionality estimation. This failure can be partially attributed to the numerical instability of LLE. However, a clear "elbow" observed around $d = 3$ suggests that $d_{in} = 3$ can still be an estimate for the original data. Moreover, we note that, different from Figures 5–7, embedding the greyscale images in the space where $d = 4, 5, 6 \ldots 10$ makes sense, and we see that the criterion value is more stable since these embeddings are not only "faithful" ($d > d_{in}$) but also "reasonable" ($D > d$).

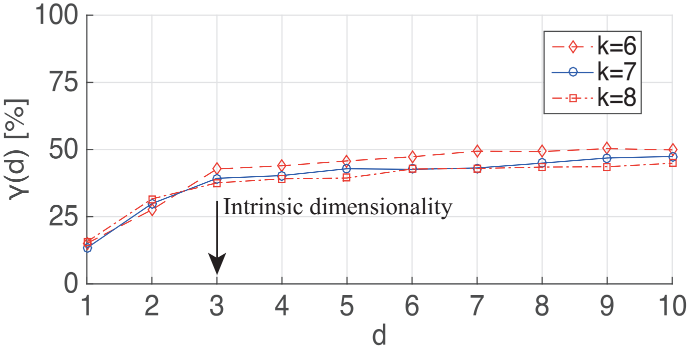

Figure 11. Neighborhood-preserving ratio for grayscale images

5. Conclusions and perspectives

In the current work, we have discussed two issues concerning LLE: the numerical instability concerning the embedded dimensionality and the neighborhood-preserving property. Theoretically, by performing LLE, one can embed the original high-dimensional data in any space (whatever its dimensionality); this former issue provoked the necessity of inferring the intrinsic dimensionality of the original data in an endeavor to find out the most compact description in the feature space. Taking into account the second issue (i.e., the evolution of neighborhoods before and after embedding), we proposed a criterion based on the ratio of preserved neighbors to estimate the intrinsic dimensionality of the original dataset. That is, we first performed LLE blindly with different imposed dimensionality d and successively depict the corresponding $d - \gamma(d)$ diagram from which the most faithful embedding in $d_{in}-$dimensional space is found.

The proposed criterion has been verified on a series of artificially generated data, covering both high- and low-dimensional one. For the illustrative examples, the proposed criterion stabilizes with an increasing dimension, and a "telltale gap"" can systematically be observed on the $d - \gamma(d)$ diagram. However, the present criterion relies on a heuristic basis; a theoretical basis is still needed.

Acknowledgements

This study was funded by National Natural Science Foundation of China (Ref No. 11620101002) and the China scholarship Council (Ref No. 201404490062). This work has also been supported by the French National Research Agency (ANR) through the research grant Micromorfing, ANR-14-CE07-0035-01.

References

[1] Balasubramanian M., Schwartz E.L., Tenenbaum J.B., de Silva V., Langford J.C.: The isomap algorithm and topological stability, *Science*, vol. 295(5552), pp. 7–7, 2002.

[2] Cox T., Cox M.: *Multidimensional Scaling*, Chapman & Hall, London, UK, 1994.

[3] Jolliffe I.: *Principal component analysis*. Wiley Online Library, 2002.

[4] Polito M., Perona P.: *Grouping and Dimensionality Reduction by Locally Linear Embedding*, 2002.

[5] Saul L.K., Roweis S.T.: An Introduction to Locally Linear Embedding. `https://cs.nyu.edu/~roweis/lle/papers/lleintroa4.pdf`, 2000.

[6] Saul L.K., Roweis S.T.: Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds, *The Journal of Machine Learning Research*, vol. 4, pp. 119–155, 2003.

[7] Saul L.K., Weinberger K.Q., Sha F., Ham J., Lee D.D.: Spectral Methods for Dimensionality Reduction, *Semisupervised learning*, pp. 293–308, 2006.

Affiliations

Liang Meng
State IJR Center of Aerospace Design and Additive Manufacturing, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an Shaanxi, 710072, P.R.China, liang.meng@nwpu.edu.cn, ORCID ID: https://orcid.org/0000-0001-7261-5918

Piotr Breitkopf
Sorbonne Universités, Université de Technologie de Compiègne, UMR UTC-CNRS 7337, Compiègne 60200, France, piotr.breitkopf@utc.fr, ORCID ID: https://orcid.org/0000-0002-3842-6520

Received: 25.04.2018
Revised: 03.07.2018
Accepted: 04.07.2018