ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

Locally Linear Embedding *versus* Isotop

John Aldo Lee, Cédric Archambeau, Michel Verleysen*

Université catholique de Louvain
Place du Levant, 3, B-1348 Louvain-la-Neuve, Belgium
{lee,archambeau,verleysen}@dice.ucl.ac.be

Abstract. Recently, a new method intended to realize conformal mappings has been published. Called Locally Linear Embedding (LLE), this method can map high-dimensional data lying on a manifold to a representation of lower dimensionality that preserves the angles. Although LLE is claimed to solve problems that are usually managed by neural networks like Kohonen's Self-Organizing Maps (SOMs), the method reduces to an elegant eigenproblem with desirable properties (no parameter tuning, no local minima, etc.). The purpose of this paper consists in comparing the capabilities of LLE with a newly developed neural method called Isotop and based on ideas like neighborhood preservation, which has been the key of the SOMs' success. To illustrate the differences between the algebraic and the neural approach, LLE and Isotop are first briefly described and then compared with well known dimensionality reduction problems.

1 Introduction

Data coming from the real world, like the outputs of sensor arrays or pixels in an image, are often difficult to understand because of their high dimensionality. Fortunately, numerous methods intended to reduce the dimensionality of a data set exist. The most known is obviously the Principal Component Analysis [4] (PCA, or Karhunen-Loeve transform). The result of this simple algebraic technique may be seen from several point of views, either as a variance preserving projection, or a minimal reconstruction error projection, or yet as distance preserving projection. In the latter case, PCA is equivalent to classical metric Multi-Dimensional Scaling [13] (MDS) under certain conditions. However, as well PCA as MDS are strictly linear models that cannot unfold nonlinear dependencies like in the first column of fig. 1. To achieve this goal, neural variants of MDS have been developed like Sammon's nonlinear mapping [10], Curvilinear Component Analysis [1, 3] and related methods [6]. One step beyond lie Kohonen's Self-Organizing Maps [5]: their power and elegance come from the fact that they do not map data by preserving pairwise distances but

*M.V. works as a senior research associate of the Belgian FNRS.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

by preserving neighborhood relations, allowing the mapping either to stretch or shrink some region of the manifold when necessary. Recently, however, neural methods have to compete with newly developed algebraic methods like Isomap [12] for distance preservation and Locally Linear Embedding [9, 11] for neighborhood preservation. An earlier paper [7] compared Isomap with CDA, while this paper confronts LLE with Isotop [8]. After a brief description of LLE in Section 2, Section 3 explains how Isotop works. Next, Section 4 shows how both algorithms apply to simple toy examples. The comparison goes on in Section 5 with the projection of a set of human faces. Finally, Section 6 draws the conclusions.

2 Locally Linear Embedding

Locally Linear Embedding (LLE) is a nonlinear dimensionality reduction based on neighborhood preservation. The mapping to a single low-dimensional coordinate system is derived from the symmetries of locally linear reconstructions.

Assuming the d-dimensional data sampled from the manifold are stored in n vectors x_i, LLE replaces each data vector by a linear combination of the k nearest other ones, leading to the reconstruction error:

$$\mathcal{E}(W) = \sum_{i=1}^{n} \|x_i - \sum_{j=1}^{n} W_{i,j} x_j\|^2 \ , \tag{1}$$

where $W_{i,j}$ are the unknowns, under a constraint of sparseness ($W_{i,j} \neq 0$ only for the k closest neighbors of each point) and an invariance constraint ($\sum_{j=1}^{n} W_{i,j} = 1$). The $W_{i,j}$ are then determined by solving a set of constrained least squares problems. The constraints ensure that the reconstructions are invariant to rotations, rescalings and translations. Once W is computed, an error function similar to $\mathcal{E}(W)$ can be written as:

$$\phi(Y) = \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{n} W_{i,j} y_j\|^2 \ , \tag{2}$$

where $W_{i,j}$ are now fixed and the unknowns are the low-dimensional coordinates y_i associated to each x_i. Under certain constraints, $\phi(Y)$ has a unique global minimum that can be computed as the solution of a sparse $n \times n$ eigenvalue problem. More details about the algorithm can be found in [11].

3 Isotop

The basic idea behind Isotop consists in overcoming some of the limitations of Kohonen's Self-Organizing Maps when they are used for nonlinear mapping. In this case, indeed, the vector quantization realized by the SOMs raises little interest and the usually rectangular shape of the map seldom suits the shape

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

of the manifold to be projected. Isotop addresses these issues in the following way. Firstly, if the dataset contains numerous points, Isotop performs a vector quantization in order to reduce their number[1]. This optional step can be achieved with simple methods, like LLoyd's algorithm or Competitive Learning, for which no neighborhood relationships intervene between the prototypes. Secondly, Isotop builds a graph structure where the nodes are the prototypes and the creation of an arc depends on the pairwise distances between the prototypes. For example, one can create arcs between one prototype and the k nearest other ones. Another solution is to link one prototype with all other ones lying closer than a given threshold ϵ. In both cases, the obtained graph structure tries to capture the neighborhood relationships in the manifold as they are underlain by the prototypes. The application of this second step yields a connected set of prototypes, comparable to the rectangular lattice of a SOM, excepted that the shape of the structure has been dynamically woven according to the shape and density of the data cloud. At this stage, the low-dimensional representation of the manifold does not exist yet. To determine it, Isotop concentrates on the graph structure. A third step is thus begun by replacing the high-dimensional coordinates of each prototype by low-dimensional ones, initialized to zero. Moreover, Isotop associates a Gaussian distribution of unit variance with each prototype, centered on it . Then, Isotop iteratively draws a point from this set of Gaussian distributions. Let g^t be the coordinates of this ramdomly generated point at time t and i the index of the prototype y_i that lies the closest from g^t. Then, Isotop updates all prototypes y_j according to the rule:

$$y_j \leftarrow y_j + \alpha^t \nu_j^t (g^t - y_j) \ , \tag{3}$$

where α^t is a time-decreasing learning rate with values taken between 1 and 0; the factor ν_j^t takes into account the previously build neighborhood relationships:

$$\nu_j^t = \exp\left(-\frac{1}{2} \frac{\delta_{i,j}^2}{(\lambda^t E_{j \in N(i)} \{\|x_i x_j\|\})^2} \right) \ , \tag{4}$$

where λ^t is a time-decreasing neighborhood width, $\delta_{i,j}$ is the graph distance computed for instance by Dijkstra's algorithm [2] and $E_{j \in N(i)}\{\|x_i x_j\|\}$ is the mean Euclidean distance between the i-th prototype and its neighbors (the set $N(i)$ gives the indices of the neighbours for the i-th prototype). Intuitively, the learning rule above unfolds the connected structure in a low-dimensional space, trying to preserve the neighborhoods. As a side effect, the mixture of Gaussian distributions evolves concurrently in order to capture the shape of the manifold. A slightly different version of Isotop (only one Gaussian distribution) is decribed with more details in [8].

[1]The quality of the quantization is not the key issue here: only its effect on the number of points to be processed afterwards is important.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

4 LLE versus Isotop: toy examples

In order to illustrate the capabilities of both LLE and Isotop, the first column of fig. 1 shows three simple manifolds taken from [11]. These are simple surfaces embedded in a three-dimensional space; theoretically, they can be mapped on a two-dimensional plane. For each manifold, thousand points are drawn (second

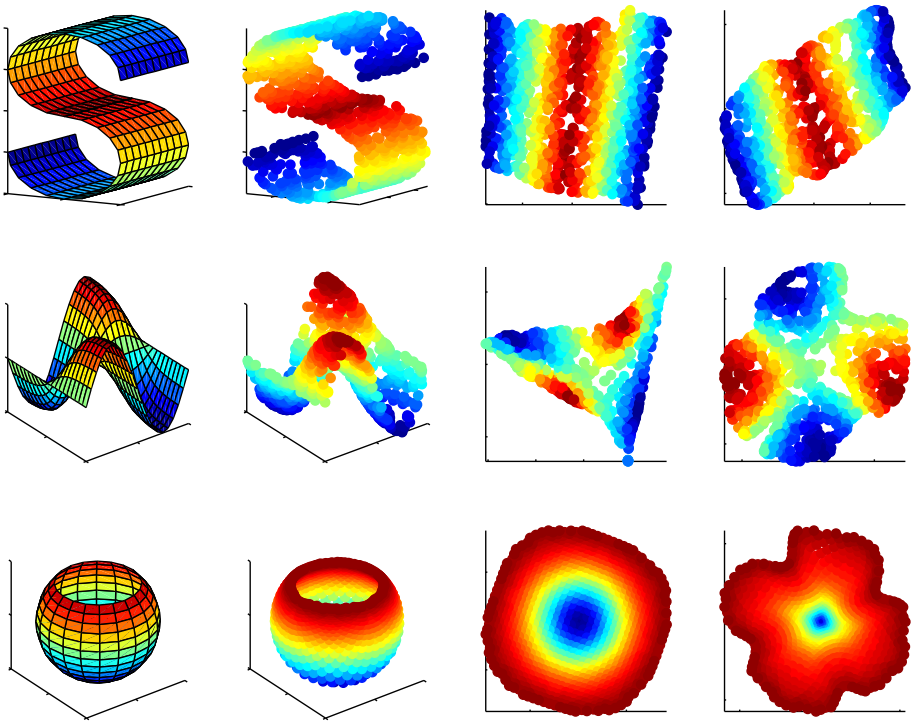

Figure 1: Projection of the S-curve, the twin peaks and the punctured sphere with Isotop and LLE: the first column shows the complete manifold, the second column shows points randomly drawn in the manifold; the third column shows the results of LLE and the fourth one the results of Isotop; color levels are used to identify the corresponding points in the third and fourth colums

column of fig. 1) before projection by LLE and Isotop (third and fourth columns of fig. 1 respectively). The k parameter of LLE was set to 8. For Isotop, no vector quantization was performed and k respectively took the values 6, 8 and 10 for the punctured sphere, the S-curve and the twin peaks. The learning rate and neighborhood width of Isotop did not need any particular care.

As it is visually evident, both methods perform rather well. The shape of the S-curve is well preserved, although LLE gives a square representation of the

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

curved rectangle. Regarding the twin peaks, the square representation of Isotop seems better than the triangular one of LLE. The situation is reversed for the punctured sphere: although both methods tends to 'square' the sphere, LLE performs better. Nevertheless, it is worth to mention that the three manifolds have been artificially generated for LLE in [11] and that the samples of the punctured sphere follow a well-specified and regular distribution. When applied to other manifolds, like the bottle shown in the left side of fig. 2, LLE totally fails, although the distribution increases towards the bottleneck in order to make the stretching of the bottleneck possible. For all acceptable value of k, LLE just delivers a linear projection of the bottle. The right part of fig. 2 shows the result of Isotop (the graph structure is still visible). In this representation, the bottleneck has been stretched and corresponds to the outer circle.

The example of the bottle shows an essential shortcoming of LLE: the appealing properties of the method (mathematical foundation, global optimizer) actually hide a model that can only project a limited class of manifolds. Once the manifold does not satisfy certain conditions (no conformal mapping exists between the manifold and a Euclidean space), the result of LLE looses all its meaning. At the price of a more intuitive foundation and some more parameterization, Isotop behaves in a smoother, more tolerant and robust way.

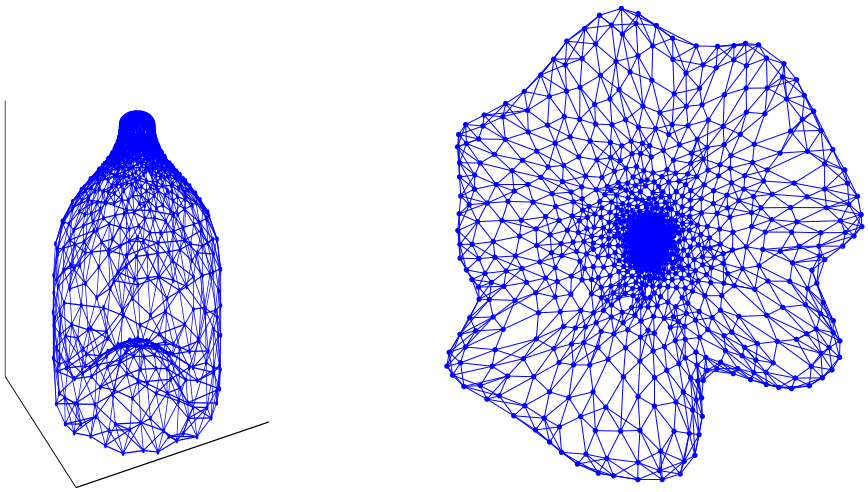

Figure 2: The bottle manifold: on the left, thousand samples are shown with the graph structure of Isotop ($k = 6$); on the right, Isotop projects the graph structure shown onto a plane, the outer circle corresponding to the stretched bottleneck

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

5 LLE versus Isotop: projection of faces

The authors of [11] have made available a dataset containing 1965 grayscale pictures of a person's face. Each picture is 20 pixels wide and 28 pixels high. Although the manifold underlain by these data has probably an intrinsic dimensionality higher than two, LLE and and Isotop can project them to a plane for visualization purpose. With $k = 12$, LLE gives the result shown in fig. 3, quite similar to the figure in [11]. Figure 3 is in fact divided in 16×16 cells that are filled with the picture corresponding to one of the points that has been projected in; empty cells are blank. With $k = 10$ and default values for the learning rate and neighborhood width, Isotop yields the result illustrated in fig. 4.

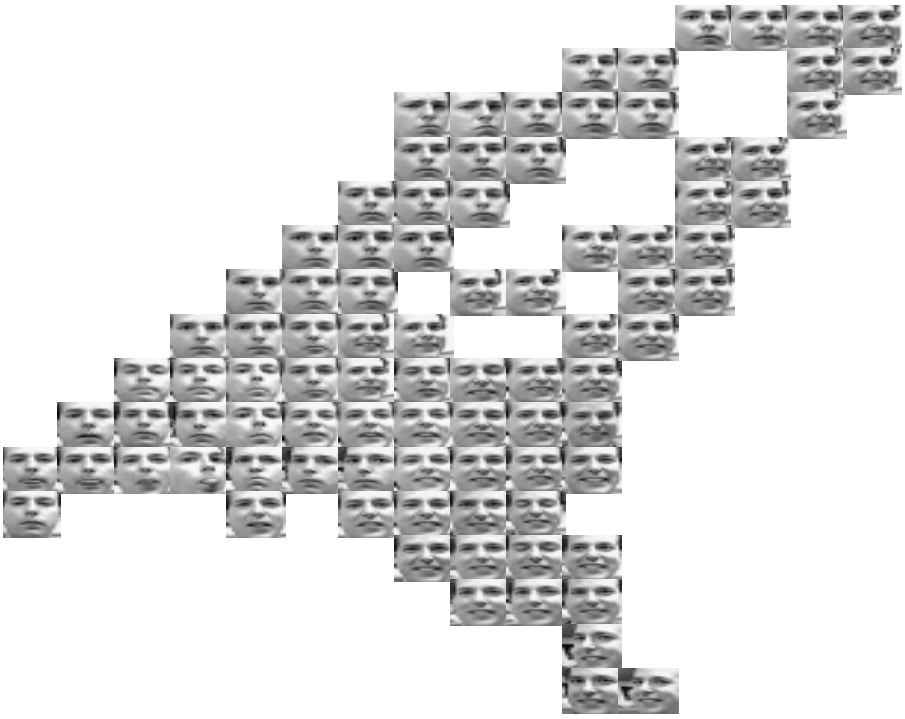

Figure 3: Projection of the face dataset by LLE

Both methods give visually satisfying results. Looking at both figures from top to bottom, one sees the head turning right to left (for LLE, this evolution is only visible on the right part of the figure). Looking at the result of Isotop from left to right, the smiling visage becomes unhappy or annoyed. Again, this effect is less visible on the result of LLE. Globally, Isotop well highlights the two dominating degrees of freedom in the dataset: the head left-right position and the visage happy-unhappy expression. Compared to the result of LLE,

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

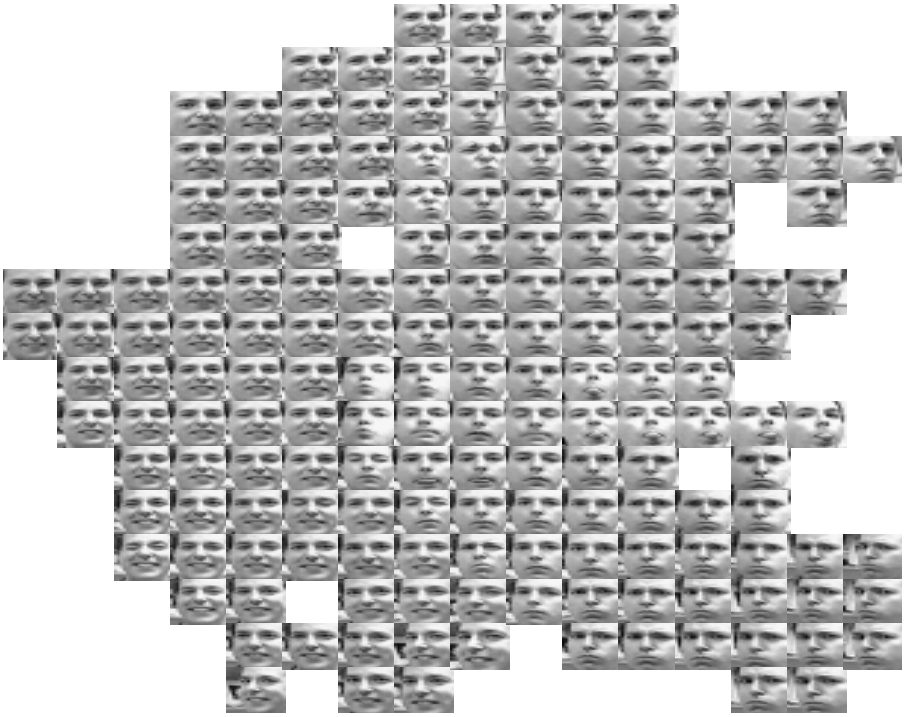

Figure 4: Projection of the face dataset by Isotop

the representation made by Isotop is distributed more uniformly, giving more importance to small details, like when the boy shows his tongue.

From a quantitative point of view, an mean organization error can be computed as follows:

$$E_{\text{org}} = \frac{1}{1965^2} \sum_{i,j=1}^{1965} \frac{|\text{rank}(x_i, x_j) - \text{rank}(y_i, y_j)|}{\text{rank}(x_i, x_j)} \quad , \tag{5}$$

where the function $\text{rank}(x_i, x_j)$ computes the rank of x_j after sorting of all distances $\|x_i - x_j\|$ measured from a fixed x_i. The denominator in the summed terms of eq. 5 gives more importance to the right ranking of the closest neighbors and E_{org} equals zero when the ranks are perfectly preserved for all pairs of points. The result of LLE in fig. 3 leads to $E_{\text{org}} = 3.4943$ while Isotop reaches the better value of 2.9179.

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 527-534

6 Conclusion

LLE and Isotop can both project nonlinear manifolds. The advantages of LLE reside in its theoretical foundations. Formulated as a simple and appealing sparse eigenvalue problem, LLE can be implemented by robust and well known algebraic procedures. However, LLE submits to these procedures sparse but very large problems that often leads to convergence failure or important numerical imprecisions. It is not uncommon that the shape of a projection made by LLE totally changes after the removal or addition of a few samples.

On the other hand, Isotop is much slower but it does not rely on generic procedures: the method works with a specifically designed neural algorithm. Isotop can indeed fall in local minima and require some care for the parameterization. As a counterpart, Isotop can map a wider class of manifolds than LLE.

References

[1] P. Demartines and J. Hérault. Curvilinear Component Analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transaction on Neural Networks*, 8(1):148–154, January 1997.

[2] E. W. Dijkstra. A note on two problems in connection with graphs. *Numerical Mathematics*, (1):269–271, 1959.

[3] J. Hérault, C. Jaussions-Picaud, and A. Guérin-Dugué. Curvilinear Component Analysis for high dimensional data representation: I. Theoretical aspects and practical use in the presence of noise. In J. Mira and J. V. Sánchez, editors, *Proceedings of IWANN'99*, volume II, pages 635–644. Springer, Alicante (Spain), June 1999.

[4] I. T. Jollife. *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[5] T. Kohonen. *Self-Organizing Maps*. Springer, Heidelberg, 2nd edition, 1995.

[6] J. A. Lee, A. Lendasse, N. Donckers, and M. Verleysen. A Robust Nonlinear Projection Method. In M. Verleysen, editor, *Proceedings of ESANN'2000, 8th European Symposium on Artificial Neural Networks*, pages 13–20. D-Facto public., Bruges (Belgium), 2000.

[7] J. A. Lee, A. Lendasse, and M. Verleysen. Curvilinear Distance Analysis versus Isomap. In M. Verleysen, editor, *Proceedings of ESANN'2002, 10th European Symposium on Artificial Neural Networks*, pages 185–192. D-Side public., Bruges (Belgium), 2002.

[8] J. A. Lee and M. Verleysen. Nonlinear Projection with the Isotop Method. In J. Dorronsoro, editor, *Proceedings of ICANN'2002, International Conference on Artificial Neural Networks*, pages 933–938. Springer, Madrid (Spain), August 2002.

[9] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[10] J. W. Sammon. A nonlinear mapping algorithm for data structure analysis. *IEEE Transactions on Computers*, CC-18(5):401–409, 1969.

[11] L. K. Saul and S. T. Roweis. Think Globally, Fit Locally: Unsupervised Learning of Nonlinear Manifolds. Technical report, University of Pennsylviana, Philadelphia, 2002.

[12] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[13] W. S. Torgerson. Multidimensional Scaling, I: Theory and Method. *Psychometrika*, 17:401–419, 1952.