Supplementary material of "H-plots for displaying non-metric dissimilarity matrices."

Irene Epifanio (epifanio@uji.es)

Departament de Matemàtiques, Universitat Jaume I, Castelló, Spain

In this file, I show more results and examples from the paper "H-plots for displaying non-metric dissimilarity matrices", together with more illustrative results on other databases: an image database of handwritten "3"s, the classical Iris data set, and the Swiss roll dataset, considered in Tenenbaum *et al.* [16] and Roweis and Saul [12]. Each database corresponds to a different section.

This work has been mainly done by using free software, R [10], and especially the MASS library [17]. The Matlab code for computing Isomap can be obtained from `http://isomap.stanford.edu`. The code is available at http://www3.uji.es/ ~epifanio/RESEARCH/hplot.rar.

1 Example 1: flights

In Fig. 1 the configuration for two dimensions for several manifold learning techniques are shown, specifically for the Locally Linear Embedding (LLE) [12], the Laplacian eigenmap (LAP) [2], the Diffusion Map (DIF) [9], and the Local Tangent Space Alignment (LTSA) [20] (the algorithms of these techniques are available from the MANIfold Learning Matlab Demo: *http//www.math.ucla.edu/~wittman/mani/*). For LLE, the distance between MA and MO is nearly the same as MA and VL, which also happens with LAP where the distance MA and MO is less than MO and SP. With DIF, VL is nearer to MO and SP than MA. However, VL and MA are at the same point with LTSA. Several σ values for DIF have been considered. For bigger values than 10, MO and SP approximate to each other, whereas MA and VL remain distant. For $\sigma = 5$, the configuration is similar to that of LLE, and for smaller values it resembles the LTSA configuration.
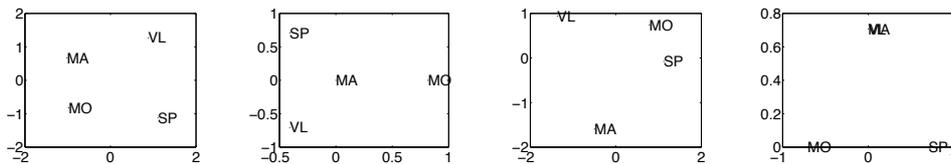

Figure 1: Scatter plot for flights: LLE (2 neighbors), LAP (2 neighbors), DIF ($\sigma = 10$, $\alpha = 1$) and LTSA (2 neighbors).

2

2 Example 3: brands

Let us suppose that the observed difference between $d_{i.}$ and $d_{j.}$ is zero for all the observed objects, except for the objects i and j. According to property 3 of the h-plot, the distance of the variables $d_{i.}$ and $d_{j.}$ in the h-plot will depend on the sample variance of their difference, and in this particular case it will depend on the size of d_{ij} in relation to the rest of the distances. This would be the 'opposite' situation to Example 1.

Let D (see Table 4) be the perceived dissimilarity matrix between five different brands of a certain kind of product on a five-point scale from 0 (indicating no dissimilarity) to 4 (maximum dissimilarity). Let us suppose that dissimilarities between brands A and B, and each of the others, are identical. Let us vary the dissimilarity between A and B from 0 to 4, which is called a, in order to see how this affects the final configuration. For $a > 2$, the triangle inequality does not hold. Note also that there are three high central objects (C, D and E).

Table 1: Dissimilarity matrix between five brands.

	A	B	C	D	E
A	0	a	1	1	1
B	a	0	1	1	1
C	1	1	0	1	2
D	1	1	1	0	4
E	1	1	2	4	0

CCs for the different methods and values of a appear in Table 2. Fig. 2 displays the configurations for my method for the different values of a. As Sammon and isoMDS need dissimilarities to be positive, except for self-

dissimilarity, I have considered $a = 0.00001$ instead of 0.

Table 2: CCs for the different methods with brands, varying a.

a	cMDS	isoMDS	Sammon	Isomap (2 neighbors)	h-plot
0.00001	0.957	0.936	0.957	0.884	0.957
1	0.946	0.940	0.951	0.884	0.934
2	0.946	0.917	0.947	0.870	0.949
3	0.933	0.901	0.934	0.873	0.935
4	0.913	0.870	0.914	0.818	0.914

My method obtains the biggest CC in all cases, except when $a = 1$, where brands A and B are represented by the same point in the h-plot. With this example, I also wanted to show that, obviously, I do not expect the h-plot to be the best for all problems. Depending on the structure of the dissimilarity matrix, better results can be obtained by other methods.
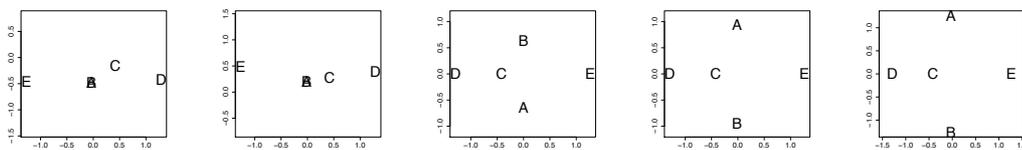

Figure 2: Scatter plot for brands with the h-plot for a from 0 (on the left) to 4 (on the right).

The configuration for all the methods when $a = 2$ is displayed in Fig. 3. For $a = 2$, all the eigenvalues in the eigendecomposition for classical scaling are positive (this indicates that D is Euclidean). Only for $a = 3$ and $a = 4$ is the fourth eigenvalue negative; for the rest of the considered values of a all the eigenvalues are positive, and therefore D is Euclidean. Note that the configuration is quite similar for the h-plot and Sammon. For cMDS, the

4

configuration is also similar, but brand C is nearer to brand E than in those methods, when their difference is 2. For isoMDS (A and C are superimposed) and Isomap (E is far from A and B), the configurations do not reflect the structure of the dissimilarity matrix so well. It seems that these two methods do not visualize the high central objects appropriately.
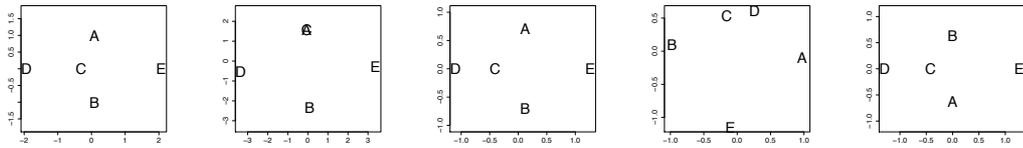

Figure 3: Scatter plot for brands: cMDS, isoMDS, Sammon, Isomap (2 neighbors) and h-plot.

Fig. 4 displays the configuration for two dimensions for the same manifold learning techniques as the previous example. For LLE, E is nearer to B than A when originally they are at the same distance, and C is more or less at the same distance from D and E. B is also nearer to E than D with LAP, and C is quite far from D. A and B are superimposed on DIF ($\sigma = 10$), while with DIF ($\sigma = 2$) a representation which is more similar to the h-plot is obtained (with $\sigma = 0.5$ C and D are superimposed). A and B are not at the same distance from C, D and E with LTSA.
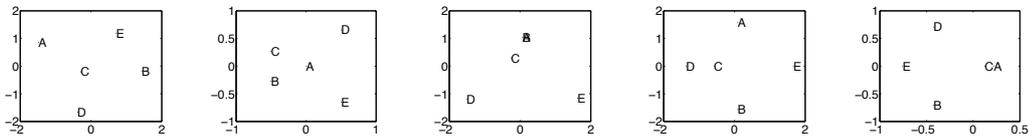

Figure 4: Scatter plot for brands: LLE (2 neighbors), LAP (2 neighbors), DIF ($\sigma = 10$, $\alpha = 1$), DIF ($\sigma = 2$, $\alpha = 1$) and LTSA (2 neighbors).

5

3   Example 4: asymmetric dataset

There are situations where d is not a distance, such as dealing with asymmetric data. My methodology can handle this kind of situation naturally, even in the case where $d_{jj} > 0$.

Let us consider a classical asymmetric dataset: the confusion percentages between 36 Morse code signals [11], which were exhaustively analyzed in Borg and Groenen [4, Sections 4.2 and 22.2]. (The dataset is available at the home page of their book: http://people.few.eur.nl/groenen/mmds/datasets/chapter 4/index.html#7). The dissimilarity matrix is a 36×72 matrix formed by the variables giving the dissimilarity (100% - rate of confusion%) from each Morse code (i.e. $d_{i.}$ denotes the dissimilarity when the i-th code is presented first), and the variables giving the dissimilarity to each Morse code (i.e. $d_{.i}$ denotes the dissimilarity when the i-th code is presented second). Fig. 5 shows the h-plot in two-dimensions, although its goodness-of-fit is not very good (78.94%). The variables $d_{i.}$ are represented in black, whereas the variables $d_{.i}$ are circled in red.

Similar observations to those appearing in Borg and Groenen [4] can be made: as we move from left to right in the figure we can see that what is changing is the temporal length of the signal (long to short), whereas as we move from the lower to the upper part of this figure the aspect of the signals that is changing is the ratio of dots to dashes (mostly dots to mostly dashes).

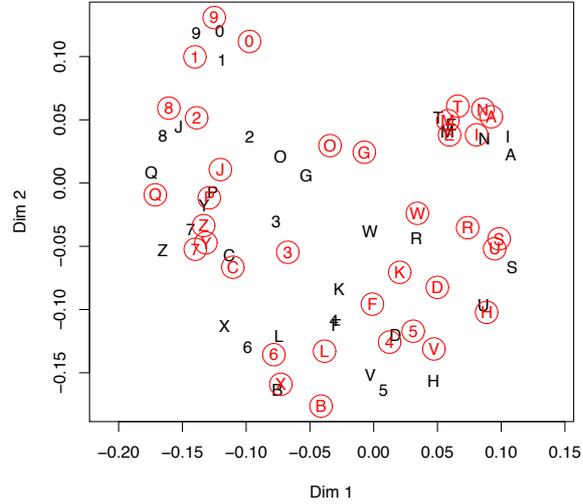According to property 3 of the h-plot, the more asymmetric codes will be

Figure 5: Scatter plot for Morse data with h-plot.

those with the greatest Euclidean distance of the variables $d_{i.}$ and $d_{.i}$ in the h-plot, which coincide with the results from the Gower decomposition [4]: H, X, and V, with $d_{.i}$ on the right of $d_{i.}$. However, Gower only considers the skew-symmetry part, not the whole matrix, so part of the information is lost. For example, S and A are very near in the Gower diagram, but they are not so near if we consider the whole matrix.

It is possible to display the symmetric part (with cMDS) together with drift vectors to model asymmetry [4, 3]. However, the representation could be misleading in some points as the symmetric and skew-symmetric parts are considered separately, not jointly as in the h-plot. For example, T, N, M are together in the Gower diagram, indicating that they are asymmetrically

7

similar, but in the cMDS configuration with drift vectors those letters are not very near because only the symmetric part is considered in the representation. With the h-plot we can see that those letters are similar.

4   Point patterns

4.1   Details of the dissimilarity

Dissimilarity between two point patterns is determined through the samples of the observed nearest neighbor distances. The *nearest neighbor distance* is defined as the distance from a point of the pattern to the closest of the other points of the pattern. Note that the point pattern is observed through a sampling window W, there are unobserved events outside the window, which can lead to mistaken estimates. This is illustrated in Fig. 6, where the observed nearest neighbor distance from the red point is bigger than the real one (to the point outside the window W). Therefore, these distances are really censored values, as was pointed out in Baddeley and Gill [1]. Let d_i denote the observed distance from the i-th point of the point pattern, s, to its nearest neighbor within the sampling window W, and c_i its distance to the complement of the sampling window W. If $c_i < d_i$, the real nearest neighbor could be outside the window and, in that case, the real and unknown nearest neighbor distance d_i' fulfills $d_i' > c_i$, and the observation is censored. In summary, the dissimilarity between two point patterns is computed by comparing the samples of the observed nearest neighbor distances, which are

really censored values.

One of the most natural ways of defining a dissimilarity measure is to borrow the ideas of two-sample hypothesis tests in survival analysis with the right censored data. The log-rank statistic [8] can be used as a judicious dissimilarity measure. Note that the usual time or duration in the context of survival analysis is replaced here by the observed distance.

In survival analysis, the time instant where an event is observed is frequently called the (observed) failure time. Here this phrase is replaced with the (observed) failure distance, i.e. the distance between a point of the pattern and its nearest neighbor, provided that these distances are smaller than the distances to the boundary of the window. In order to construct the log rank test, a sequence of 2×2 tables is built over the distances (one for each failure distance, d_j), where the risk set at that distance is classified into a 2×2 table, according to the group and event status. Let us present some notation: r_j stands for the size of the risk set in the interval $[d_{j-1}, d_j)$ and f_j the number of failures. The size of the risk and failure sets for each group can also be defined. In this way, for the first group they are denoted by r_{1j} and f_{1j}, respectively. Differences between the observed and the expected events (conditioning on the margins of the 2×2 table) in the first group are totalled over distances and squared in order to calculate the numerator of the log-rank statistic:

$$U = \sum_j (f_{1j} - g_{1j}), \qquad (1)$$

9

where g_{1j} stands for the expected number of failures in the first group, g_{1j} $= f_j * r_{1j}/r_j$. The denominator is the sum of variances for the number of events in the first group within each 2×2 table, which is obtained by using the hyper-geometric distribution:

$$Var(U) = \sum_j r_{1j} * f_j/r_j(1 - f_j/r_j)(r_j - r_{1j})/(r_j - 1). \qquad (2)$$

The code for computing the dissimilarity between two point patterns using the log-rank statistic with the nearest neighbor distance samples is available, as previously mentioned.

4.2 Details on the generation of the experiments

Three experiments have been considered. The simulated data is meant to look similar to the corneal endothelium data. Different Strauss processes have been chosen. (A Strauss process is a Markov point process where the conditional intensity of an event at a point x, given the realization of the process in the remainder of any set B, is αc^{s_x}, where s_x is the number of events within a distance r from x. They are called the neighbors of x and parameter c, with $0 \leq c \leq 1$ describes the interaction between neighbors, while parameter α controls the intensity of the process.)

The first experiment considers different Strauss processes in a unit window i.e. different degrees of regularity, all of them with $\alpha = 120$ and $c = 0$, and the three groups correspond to three values of r (i.e. three interaction distances):

10

0.01, 0.04 and 0.07. One hundred patterns have been simulated from each group. Figure 7 displays one of the 100 point patterns generated for each group. Note that I compute the dissimilarity between these point patterns, not within them.

In the second experiment the same processes are considered but the window is changed. One half of these patterns are now generated in the unit window $W_1 = [0, 1] \times [0, 1]$ with the other half in the window $W_2 = [0, 0.5] \times [0, 0.5]$, i.e. 50 point patterns in W_1 and 50 in W_2, for each group. This situation occurs frequently in my motivating example, where the window size is variable, as corneal endothelia are of variable sizes.

The third experiment is also similar to the first one, but instead of changing the interaction radius I will change its strength. Different Strauss processes with $\alpha = 120$ define every group with a fixed $r = 0.07$, but with different values of c: 0, 0.5 and 1. When parameter c increases, the probability of finding pairs of points at a distance of less than r increases.

4.3   Simulated examples

The three experiments that I have done share the same experimental setup. In each of them, three different groups of point patterns are considered. Each group is composed of 100 point patterns generated from the same model with the same parameters. Obviously, in this way I know the true partition.

For the dissimilarity matrices of the previous data (I have three matrices of 300×300), I have computed the CC shown in Table 3. The goodness-of-fit

for my method with these experiments appears in Table 4, giving an excellent fit for two dimensions. In all cases the highest CC is achieved by the h-plot but, as before, pictures provide a better comparison of the results.

As the true partition is known, the real groups are represented by different points. The first group appears as blue squares, the second one as red circles and the third one as green triangles.

Results for the first experiment are shown in Fig. 8, for the established methods and for my method, both using the original dissimilarities and their ranks (the original dissimilarities are replaced with their sample ranks). I get a clear separation of the groups with my method in both dimensions, as opposed to the other methods which separate the clusters with varying success with respect to the first dimension. Note that some blue squares are in the red circle group for all the methods except the h-plot.

Table 3: CCs for the different methods with simulated point patterns.

Experiment	cMDS	isoMDS	Sammon	Isomap (25 neighbors)	h-plot
First	0.965	0.971	0.967	0.929	0.974
Second	0.875	0.875	0.791	0.283	0.879
Third	0.95	0.956	0.955	0.891	0.962

Table 4: Goodness-of-fit of my method for one and two dimensions with simulated point patterns.

Experiment	One dim.	Two dim.
First	97.573	99.99
Second	88.895	99.7
Third	97.99	99.996

Table 5: Average of the original dissimilarities for the subgroups of the second experiment.

Average dissimilarities	$1\ W_1$	$1\ W_2$	$2\ W_1$	$2\ W_2$	$3\ W_1$	$3\ W_2$
$1\ W_1$	1.7	2.02	20.38	6.44	61.71	13.71
$1\ W_2$	2.02	1.82	15.32	6.33	47.08	12.3
$2\ W_1$	20.38	15.32	1.98	1.9	27.21	6.15
$2\ W_2$	6.44	6.33	1.9	1.83	20.17	5.61
$3\ W_1$	61.71	47.08	27.21	20.17	2.25	2.22
$3\ W_2$	13.71	12.3	6.15	5.61	2.22	1.76

Results for established methods are displayed in Figs 9 and 11 for experiments 2 and 3, respectively, while results for my method are shown in Figs 10 and 12, both using the original dissimilarities and their ranks. The lightest colors in Figs. 9 and 10 indicate that the group was generated in the window $W_2 = [0, 0.5] \times [0, 0.5]$.

We can appreciate a clear separation of the groups with my method in the third experiment. However, this separation does not appear in the majority of methods in the second experiment. An analysis of the second experiment dissimilarities reveals the reason for the point arrangement in the methods. Table 5 shows the average of the dissimilarities for the three groups according to the window where they were generated: $W_1 = [0, 1] \times [0, 1]$ or $W_2 = [0, 0.5] \times [0, 0.5]$. The precision was different for those different window sizes. If I use the rank of the dissimilarities with the h-plot for discovering clusters, a much better arrangement is obtained (see Fig. 10 (b)).

5   Images of Handwritten "3"s

This dataset is available at `http://www-stat.standford.edu/ElemStatLearn`, the website for the book [6]. It consists of normalized handwritten digits, automatically scanned from envelopes by the U.S. Postal Service. The original scanned digits have been normalized to be of approximately the same size and orientation, resulting in 16 x 16 grayscale images [7]. I have chosen the 824 available images of the digit 3.

In order to apply the different methods of multidimensional scaling, I have computed a well-known objective measure between each pair of images [15]: the mean square error (MSE). Therefore, the element d_{ij} of the dissimilarity matrix takes the following form:

$$d_{ij} = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} (f_i(x,y) - f_j(x,y))^2$$

where M and N are the horizontal and vertical dimensions of the images, and f_i and f_j stand for the i-th and j-th images, giving the brightness values for each pixel. Although MSE is related to the Euclidean distance, which was used for determining the distance between two images in Tenenbaum *et al.* [16], I have preferred MSE because it is more common in the image processing field. This dissimilarity is not a metric, since triangle inequality is violated. In this example, the smallest 667 eigenvalues of the matrix B in classical multidimensional scaling (B is defined in the paper) are negative.

Table 6 shows the CC for the different methods. In all cases, it is near

Table 6: CCs for the different methods with Handwritten "3"s.

cMDS	isoMDS	Sammon	Isomap	h-plot
0.940	0.958	0.940	0.925	0.916

one. On the other hand, the goodness-of-fit for my method for one and two dimensions is 77.12% and 92.01%, respectively. However, what are really meaningful are the pictures themselves, since they show the reduced configurations in detail. The scatter diagrams for isoMDS (the method with the highest CC) and Isomap (considering the 25 nearest neighbors) appear in Fig. 13, while my method is displayed in Fig. 14. A sample of the images is superimposed onto the 2D projection of all input images. The red point indicates the point representing the image next to it. This representation makes it easier to show the data's underlying geometric structure. For my method, the upper-left region seems to represent "slim" threes, and looking from left to right, we observe "3"s with longer upper and lower tails. The roundest "3"s are in the bottom-left part. It is difficult, however, to distinguish visually any clear structure in the configuration given by isoMDS. The roundest "3"s are in the central part, but the periphery is not so easy to interpret. With regard to Isomap configuration, the slimmest "3"s are in the bottom to center right-hand part, and the majority of "3"s with longer upper and lower tails are in the left and upper part, although there are also some in the central part, together with the roundest "3"s. It seems that my representation provides a more perceptually meaningful structure.

15

Table 7: CCs for the different methods with the Iris data set.

cMDS	isoMDS	Sammon	Isomap	h-plot
0.999	0.999	0.999	0.995	0.988

6　Iris data set

The Iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica. This is a classical dataset used in many articles [19]. I have considered the Euclidean distance in order to build the dissimilarity matrix. In this way, the Iris setosa class is well-separated from the other two classes which are not as easily separated from each other.

Table 7 shows the CC for the different methods. Now they are nearer to one than in the previous example. Furthermore, the goodness-of-fit for my method for one dimension is very high, 98.58%, providing 99.99% for two dimensions. The scatter diagrams for the established methods appear in Fig. 15, where each symbol represents a different species (blue square for Iris setosa, red circle for versicolor and green triangle for virginica). My method is shown in Fig. 16 (a).

There is a clear separation of the setosa species from the others, in my method and Isomap being a tighter cluster than in the other representations.

In my representation, the phenomenon called horseshoe or Guttman effect appears [14]. The solution has a U shape, which would indicate that

16

Table 8: Means for each group and variable of Iris data.

Group	Sepal Length	Sepal Width	Petal Length	Petal Width
setosa	5.006	3.428	1.462	0.246
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026

the data is basically one dimensional. In this case, the one-dimensional co-ordinate could refer to the size of the flowers. Direct inspection reveals that setosa is physically much smaller than the other two species, and versicolor is somewhat smaller than virginica (see Table 8). The intrinsic dimensionality of the data estimated by the 'elbow' of the residual variances plot as suggested in Tenenbaum *et al.* [16] is, however, four (see Fig. 17).

My method applied to the dissimilarity ranks (original dissimilarities are replaced with their sample ranks) is shown in Fig. 16 (b). It seems that now the species versicolor and virginica are somewhat more separated (there is a small gap at (5000,-1000)).

7   Swiss roll

The Swiss roll dataset is a randomly sampled plane rolled up into a spiral. It has appeared in several studies [16, 12]. I have considered the first 1000 points of the Swiss roll from the package Isomap for Matlab `http://Isomap.stanford.edu`, and each component has been standardized. Afterwards, the dissimilarity matrix is computed with the Euclidean distance (a metric measure).

Figure 18 shows these points. Figs 19, 20 and 21 show the configurations

17

cMDS	isoMDS	Sammon	Isomap (7)	LLE (7)	LLE (25)	HLLE (7)
0.965	0.970	0.969	0.843	0.863	0.956	0.922
LAP (7)	LAP (25)	LTSA (7)	CA	DIF (σ=10)	DIF (σ=0.5)	h-plot
0.883	0.910	0.922	0.963	0.965	0.954	0.962

Table 9: CCs for the different methods with Swiss roll.

for different multidimensional scaling methods (using the same colors as in Fig. 18 for representing the same points): cMDS, isoMDS, Sammon, together with several manifold learning techniques from the MANIfold Learning Matlab Demo. These methods are: Isomap, LLE, the Hessian eigenmaps (HLLE) [5], LAP, DIF, and LTSA. These methods are controlled by one or several parameters, and very different results can be obtained depending on the values of the parameters [13]. Correspondence analysis (CA) as in Weller and Romney [18] is also considered. The dissimilarity data must first be transformed into similarity data. The same strategy as in Weller and Romney [18] is followed: subtracting them from a number larger than the largest value in the dissimilarity matrix, 5 in this example. Note that when using CA for non-frequency data, some precautions should be born in mind [18, pp. 70-76]. The five precautions discussed by Weller and Romney [18, pp. 70-76] are: first, inferential tests are not valid; second, the data must be in the form of similarities; third, it is critical that the diagonal values have a large positive value; fourth, all values must be positive (zeros are acceptable) and all row and column totals must be greater than zero; five, in sparse matrices be alert to "defects" in data, such as two or more disjoint sets, which have to be analyzed separately.

Table 9 shows the CC for the different methods. This example reflects the different approaches considered by classical multidimensional scaling methods and the recent manifold learning techniques developed for non-linear dimensionality reduction. Classical multidimensional scaling techniques try to preserve pairwise distances in order to study the proximity structure between pairs of objects, whereas for the new nonlinear dimension reduction methods such as Isomap, LLE, etc., data are assumed to lie on a nonlinear manifold [13], and in this specific example, their goal is to "flatten" the Swiss roll, thus eliminating the dimensions that are used up by curvature. This is the reason why blue points are represented in two separate groups (see, for example, points with coordinates -6 and 2 in the first dimension for the Isomap in Fig. 19), despite their closeness (see Fig. 18). This situation does not occur in the classical multidimensional scaling methods and CA, but in these representations pink points are intermingled with green and brown points when they are not neighbors (see Fig. 18). Neither of the situations previously mentioned for the classical and the new methods such as Isomap appear with the h-plot. The configuration obtained with the h-plot resembles a spiral, being visually similar to that obtained with the Diffusion map ($\sigma=0.5$, $\alpha=1$). Note also that the configuration obtained with the Diffusion map ($\sigma=10$, $\alpha=1$) is quite different, showing that parameter selection is a critical and key factor, as previously mentioned.

19

References

[1] Baddeley, A. and Gill, R. D. (1997). Kaplan-Meier estimators of distance distributions for spatial point processes. *The Annals of Statistics*, **25**(1), 263–292.

[2] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, **15**(6), 1373–1396.

[3] Borg, I. and Groenen, P. (1995). Asymmetries in multidimensional scaling. In *SoftStat '95: Advances in Statistical Software*, pages 31–35.

[4] Borg, I. and Groenen, P. (1997). *Modern multidimensional scaling theory and applications*. Springer.

[5] Donoho, D. L. and Grimes, C. (2003). Hessian eigenmaps: locally linear embedding techniques for high dimensional data. *Proc. of National Academy of Sciences*, **100**(10), 5591–5596.

[6] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning. Data mining, inference and prediction*. Springer-Verlag.

[7] LeCun, Y., Boser, B., Denker, J. S., Howard, R. E., Habbard, W., Jackel, L. D., and Henderson, D. (1990). *Advances in neural information processing systems 2*, chapter Handwritten digit recognition with a back-propagation network, pages 396–404. Morgan Kaufmann Publishers Inc.

[8] Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**, 719–748.

[9] Nadler, B., Lafon, S., Coifman, R. R., and Kevrekidis, I. G. (2006). Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, **21**, 113–127.

[10] R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. ISBN 3-900051-07-0.

[11] Rothkopf, E. (1957). A mesure of stimulus similarity and errors in some paired-associated learning. *Journal of Experimental Psychology*, **53**, 94–101.

[12] Roweis, S. T. and Saul, L. K. (2000). Linear embedding nonlinear dimensionality reduction by locally. *Science*, **290**(5500), 2323–2326.

[13] Saul, L. K. and Roweis, S. T. (2003). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, **4**, 119–155.

[14] Seber, G. A. F. (1984). *Multivariate observations*. John Wiley.

[15] Shi, Y. and Sun, H. (2000). *Image Compression for Multimedia Engineering: fundamentals, algorithms and standards*. CRC Press, Boca Raton.

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

[17] Venables, W. N. and Ripley, B. D. (2002). *Modern applied statistics with S-plus*. Springer.

[18] Weller, S. C. and Romney, A. K. (1990). *Metric Scaling: Correspondence analysis*. Sage, Newbury Park, CA.

[19] Yang, L. (2004). Distance-preserving projection of high-dimensional data for nonlinear dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(9), 1243–1246.

[20] Zhang, Z. and Zha, H. (2004). Principal manifolds and nonlinear dimension reduction via tangent space alignment. *SIAM Journal of Scientific Computing*, **26**(1), 313–318.
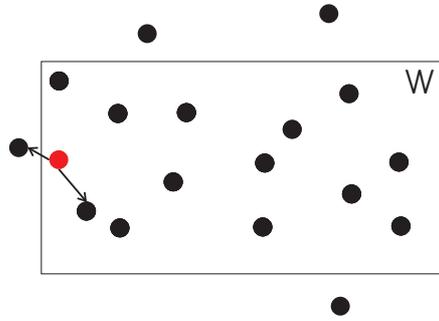
Figure 6: Edge effect illustration.

(a)   (b)   (c)
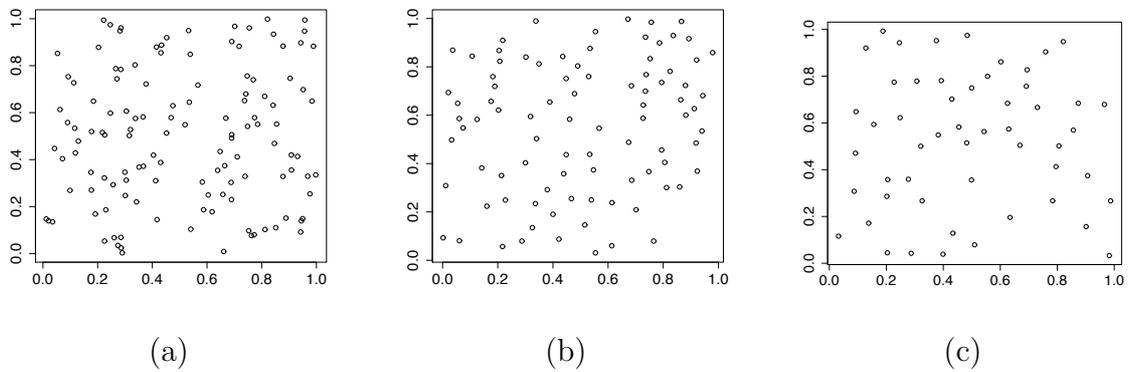
Figure 7: First experiment: different Strauss point patterns with $c=0$, $\alpha=120$ and (a) $r=0.01$, (b) $r=0.04$ and (c) $r=0.07$.
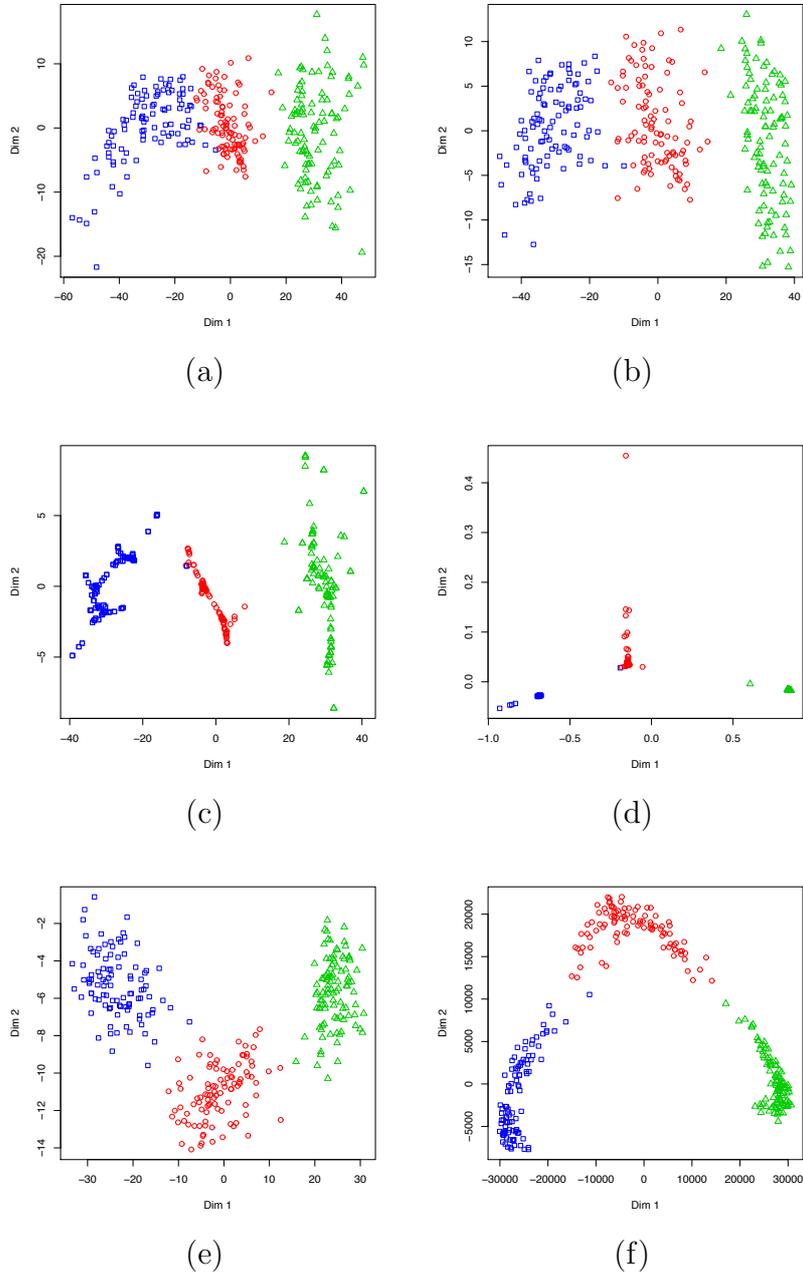
Figure 8: First experiment: (a) cMDS, (b) isoMDS, (c) Sammon, (d) Isomap (25 neighbors), and h-plot using: (e) the original dissimilarities, and (f) the dissimilarity ranks.

24
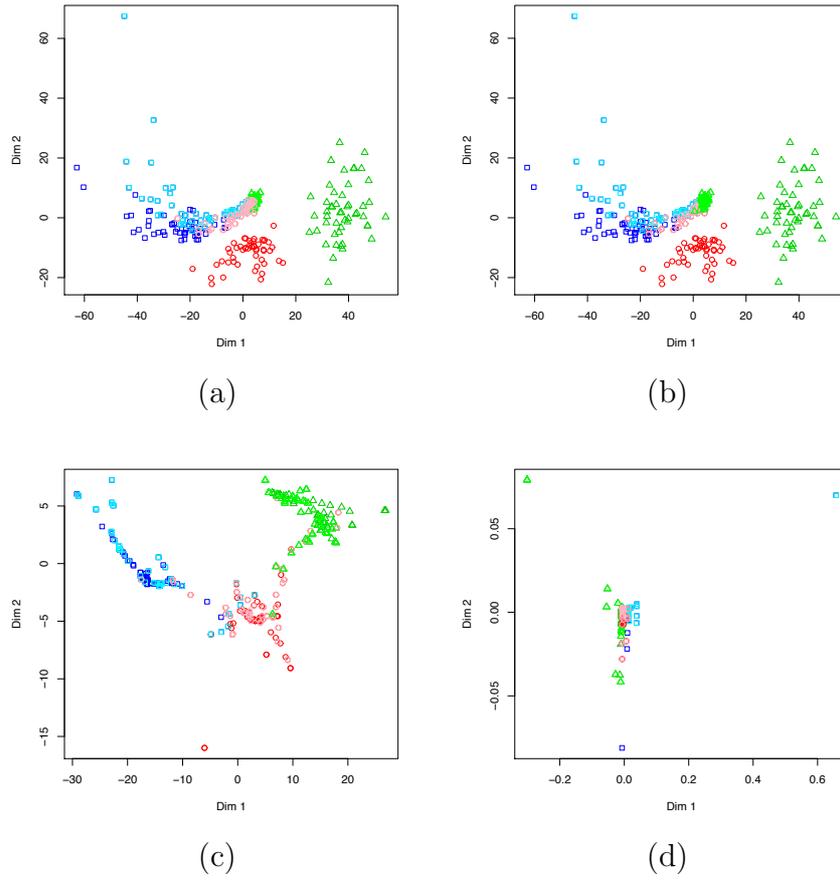
Figure 9: Second experiment: (a) cMDS, (b) isoMDS, (c) Sammon, and (d) Isomap (25 neighbors).
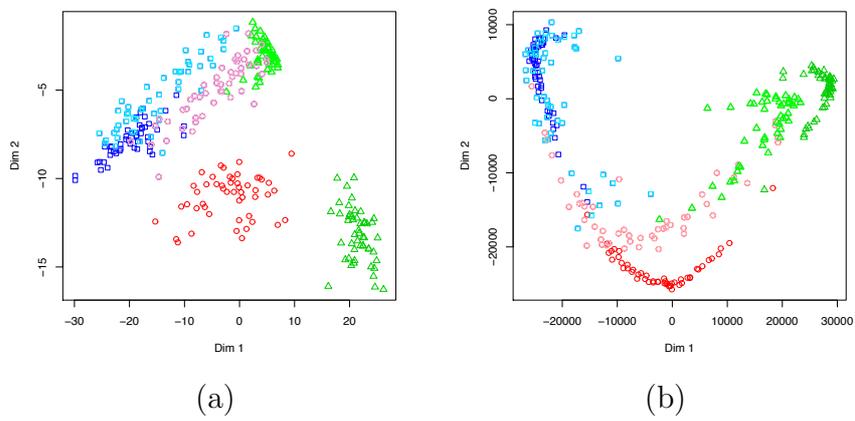
Figure 10: Second experiment with the h-plot using: (a) the original dissimilarities, and (b) the dissimilarity ranks.
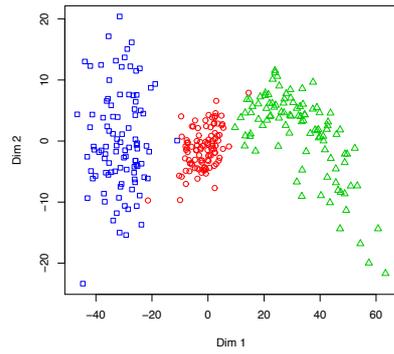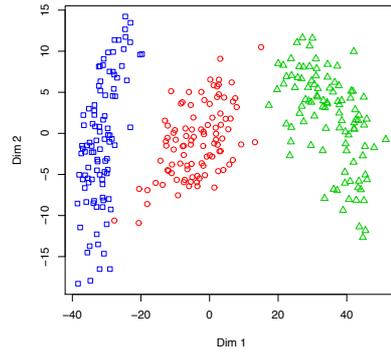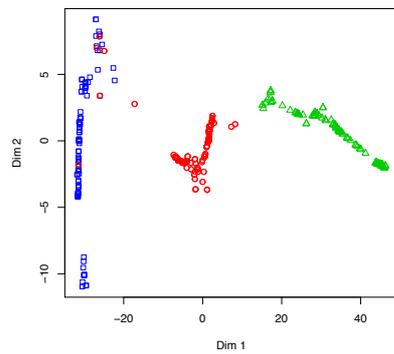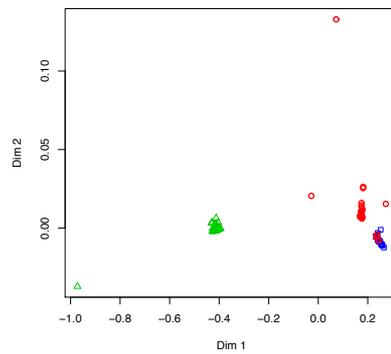
Figure 11: Third experiment: (a) cMDS, (b) isoMDS, (c) Sammon and (d) Isomap (25 neighbors).
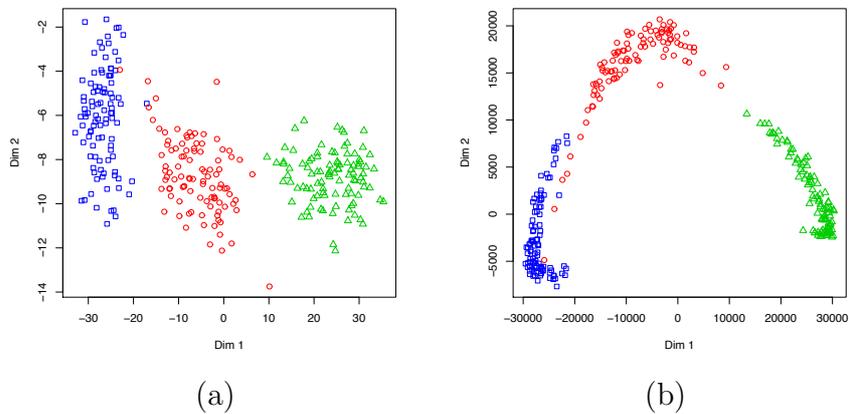
Figure 12: Third experiment with the h-plot using: (a) the original dissimilarities, and (b) the dissimilarity ranks.
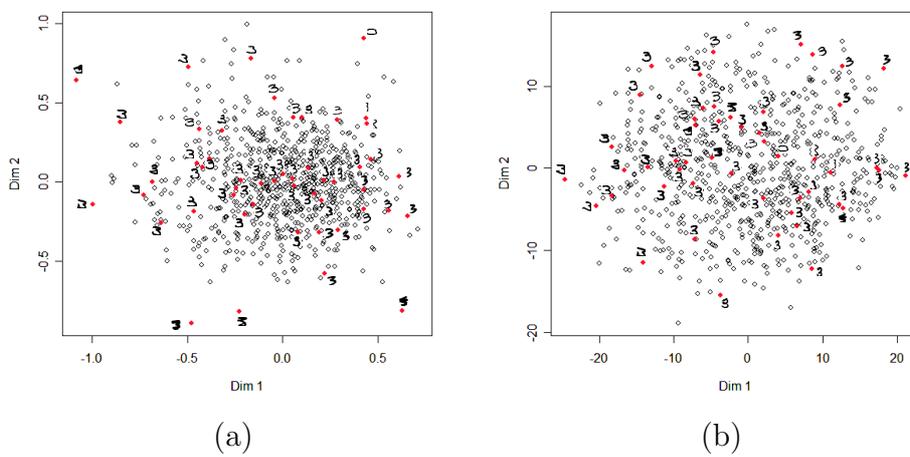
Figure 13: Scatter plot for handwritten "3"s with: (a) isoMDS and (b) Isomap.
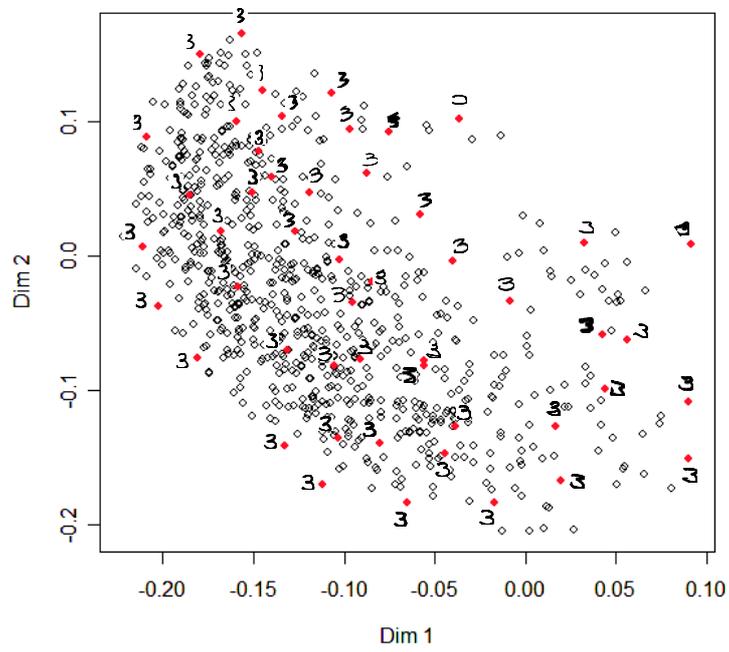
Figure 14: Scatter plot for handwritten "3"s with the h-plot.
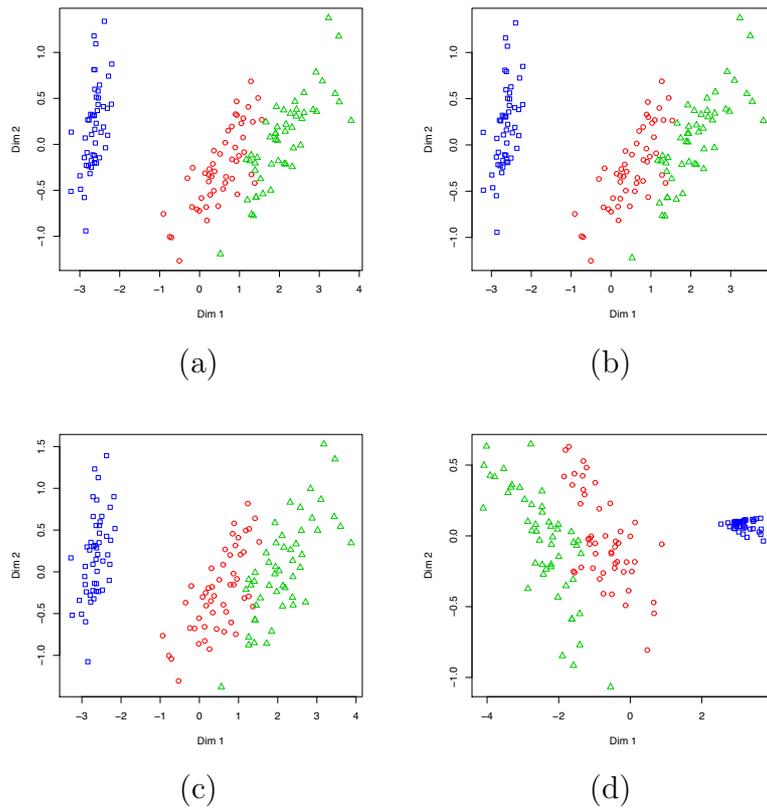
Figure 15: Scatter plot for Iris data set with: (a) cMDS, (b) isoMDS, (c) Sammon and (d) Isomap (considering the 25 nearest neighbors as in Yang [19]).

(a)   (b)
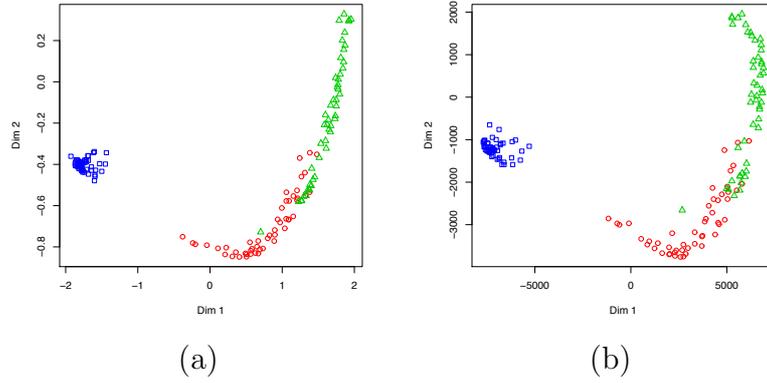
Figure 16: Scatter plot for Iris data set with the h-plot using: (a) the original distances, and (b) the distance ranks.
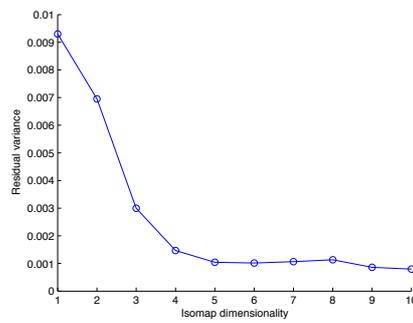
Figure 17: Residual variance of Isomap versus Dimension.

31
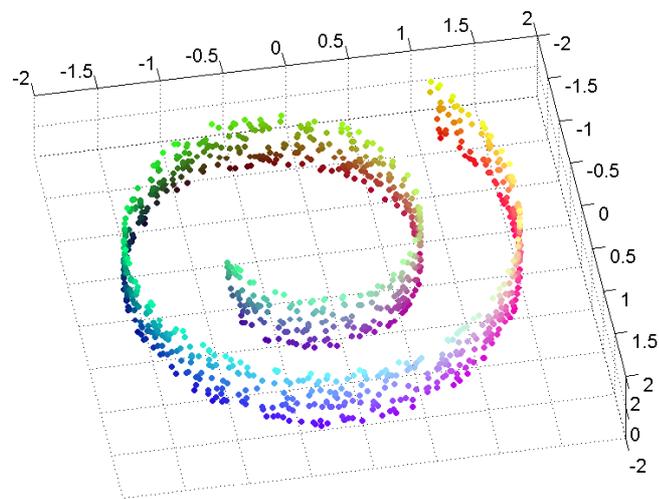
Figure 18: Swiss roll.
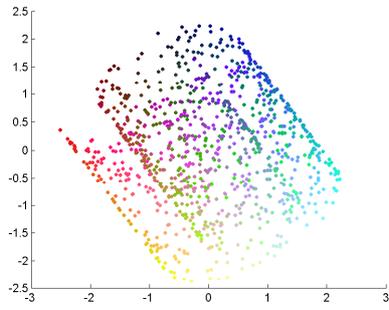
cMDS
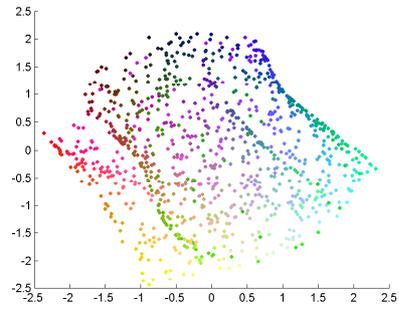
isoMDS
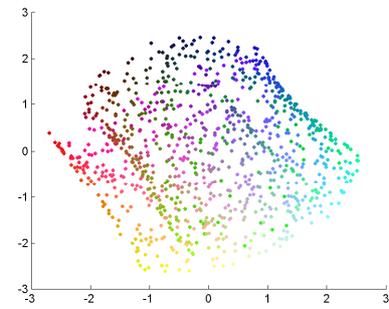
Sammon
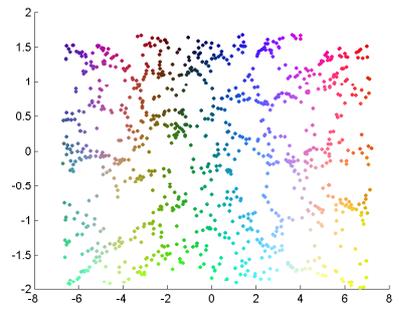
Isomap (7 neighbors as in Tenenbaum *et al.* [16])
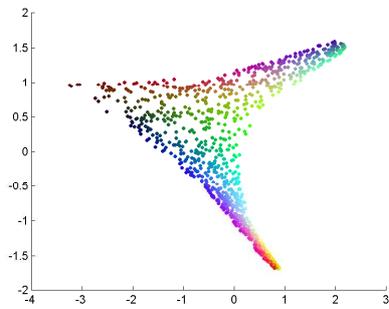
LLE (7 neighbors)
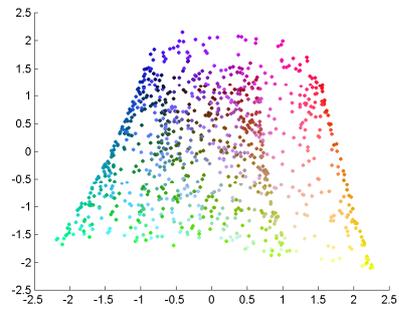
LLE (25 neighbors)

Figure 19: Scatter plots for Swiss roll (part I).
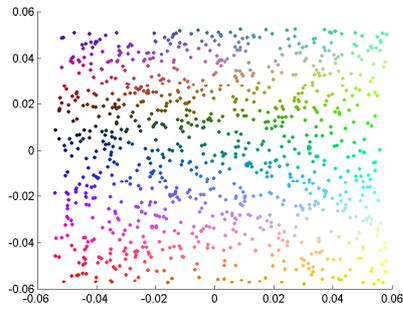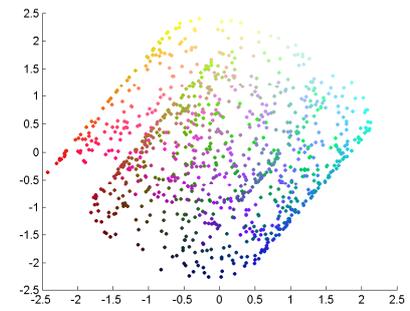
Hessian LLE (6 neighbors)
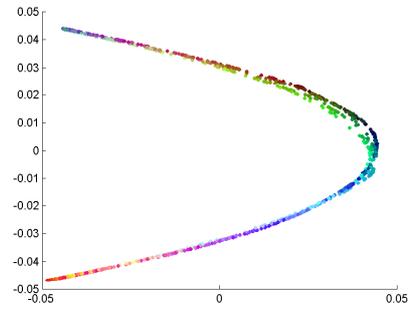
Laplacian (7 neighbors)
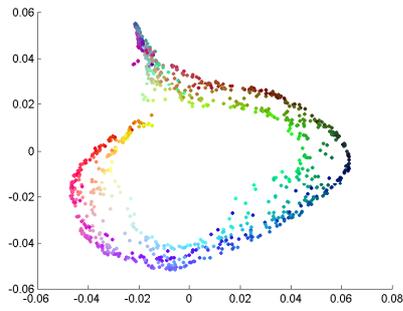
Laplacian (25 neighbors)
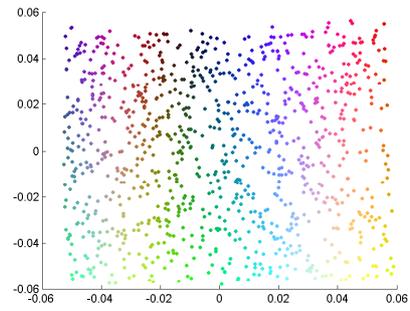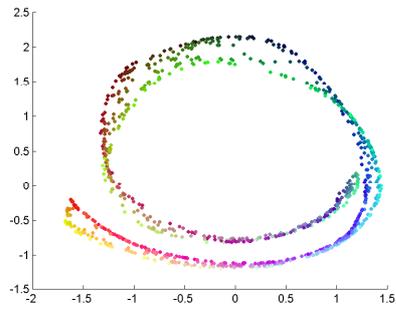
LTSA (7 neighbors)
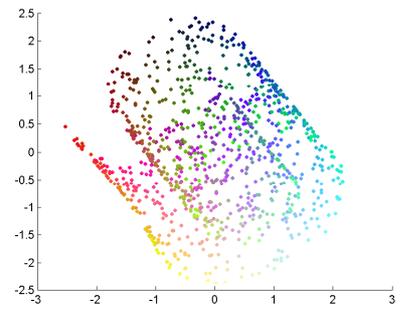
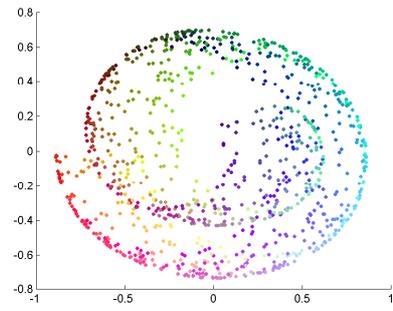Correspondence analysis of similarities

Diffusion map ($\sigma=10$, $\alpha=1$)

Figure 20: Scatter plots for Swiss roll (part II).

Diffusion map (σ=0.5, α=1)   h-plot

Figure 21: Scatter plots for Swiss roll (part III).