# H-plots for displaying non-metric dissimilarity matrices

Irene Epifanio (epifanio@uji.es)

Departament de Matemàtiques, Universitat Jaume I, Castelló, Spain

**Abstract**

Non-metric pairwise data with violations of symmetry, reflexivity or triangle inequality appear in fields such as image matching, web mining or cognitive psychology. When data are inherently non-metric, we should not enforce metricity as real information could be lost. The multidimensional scaling problem is addressed from a new perspective. I propose a method based on the h-plot, which naturally handles asymmetric proximity data. Pairwise proximities between the objects are defined, though I do not embed these objects, but rather the variables that give the proximity to or from each object. The method is very simple to implement. The representation goodness can be easily assessed. The methodology is illustrated through several small examples and applied to the analysis of digital images of human corneal endothelia. Comparisons with well-known methods show its good be-

havior, especially with non-metric pairwise data, which motivate my methodology. Other databases and methods are analyzed in the supplementary material.

Keywords: Multidimensional Scaling, Proximity Data, non-Euclidean pairwise data, Embedding, Visualization.

# 1 Introduction

Multidimensional scaling is a classical problem in many fields. However, there has been a resurgence of interest in the dimensionality reduction problem, as evidenced by the surge of publications in this field over the last few years [4, 22, 24, 25]. This could be due to the acquisition of large volumes of high-dimensional data produced by technological advances, and the fact that in some fields such as bioinformatics, features are unavailable [10, Ch.18], and only proximity information between pairs of objects is available. In some fields such as image matching, text or web mining or cognitive psychology, pairwise data are non-metric and the dissimilarity matrix does not satisfy the mathematical requirements of a metric function (reflexivity, definiteness, symmetry, triangle inequality). In these cases, non-metricity is not due to noisy measurements, but due to data being inherently non-metric. In these cases, we should not enforce metricity (for example, adding a constant to the non-diagonal dissimilarities), as real information could be lost [17]. The great majority of well-established machine learning methods have been formulated

for metric data only, and often the metricity violation is not taken into account, transforming the dissimilarities (for instance, by symmetrization) or simply omitting the negative eigenvalues as in classical scaling. I propose a method based on the h-plot, which can take into account the non-metricity.

Many methods for multidimensional scaling carry out data projection by iteratively minimizing some kind of distance-based error measure. In this paper, I propose an alternative for the previous methods, with the advantage of having an explicit solution in terms of eigenvectors. Furthermore, the representation goodness can be easily assessed and it can be applied to asymmetric proximity data in a natural way. Additionally, I show how my basic method can be modified if clustering and pattern detection is a priority.

Applications of multidimensional scaling are numerous in many fields (psychology, marketing, ecology, molecular biology, computational chemistry, social networks, graph layout or music) [11]. Here, I consider the problem of displaying dissimilarity matrices when the observations are images, i.e. $d_{ij}$ gives the dissimilarity between the images $i$ and $j$.

My motivating problem is concerned with the analysis of digital images of human corneal endothelia. The endothelial cell population decreases with age or following stressful situations such as cataract surgery, corneal transplantation or the implantation of intra-ocular lenses. When endothelial loss occurs, the endothelial response is an enlargement and sliding of the existing cells to cover the area previously occupied by the lost cells. As a result, the cells lose their hexagonal appearance. Figure 1 shows an example of a human

corneal endothelium.

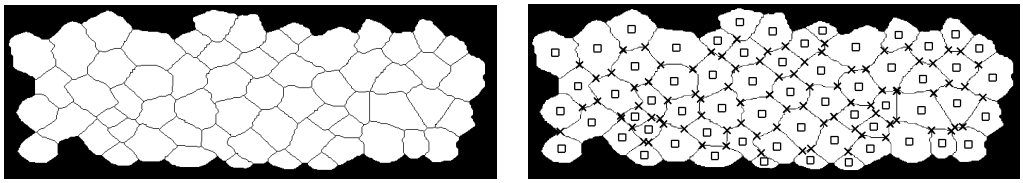

Figure 1: A human corneal endothelium and the corresponding centroids (squares) and triple points (crosses).

In Ayala *et al.* [1], a methodology for finding groups corresponding to different morphologies of the corneal endothelia was presented. Their basic idea was to associate two point patterns to a given image. Under the ideal model, the normal endothelium is expected to be a regular tessellation, which can be characterized by the centroids of the cells or by those points corresponding to the apical intersections, which are triple points (those points where three different cells meet). In this theoretical model, both point sets would be symmetrically located regular grids. The similarity between images was reformulated as a similarity between the corresponding point patterns.

In this paper, the graphical exploration of those (non-metric) dissimilarity matrices is carried out by using the method proposed based on h-plots. My method is introduced in Section 2, and its benefits are illustrated through two examples. In Section 3 my method is compared with other multidimensional scaling methods through the dissimilarity matrix defined for the images of human corneal endothelia. More comparisons with other methods and other databases are available in the supplementary material. The code for reproducing the examples is available at http://www3.uji.es/∼epifanio/RESEARCH/

hplot.rar. Finally, some conclusions are given in Section 4.

## 2 Methodology

The biplot is a very useful tool for graphically observing the structure of large matrices [8, 9]. The biplot provides a simultaneous summary of the relationships among the observations and the relationships among the variables. Corsten and Gabriel [5] used it for comparing variance-covariance matrices, where only variables were represented in the plot called an h-plot.

There are many variations in biplots, but the most widely-used one is considered here, as introduced in Seber [23]. In the ensuing discussion, I describe how the h-plot is computed.

Let $X$ be a $n \times m$ data matrix (in which each column corresponds to a variable and each row corresponds to an object or individual), and $S$ the unbiased estimate of the variance-covariance matrix. If we are interested in the h-plot in two dimensions, the two largest eigenvalues $\lambda_1$ and $\lambda_2$, with corresponding unit eigenvectors $q_1$ and $q_2$ of $S$, can be found. Note that the eigenvalues are non-negative, since the covariance matrix $S$ is always positive semi-definite [15]. I then build (and represent) the matrix

$$H_2 = (\sqrt{\lambda_1}q_1, \sqrt{\lambda_2}q_2).$$

Rows $h_j$ of matrix $H_2$ have, approximately, the following properties:

1. The sample covariance $s_{ji}$ between variables $j$ and $i$ is $h_j' h_i$, where $'$ indicates the transposition. Hence, the sample variances $s_{jj}$ are $||h_j||^2$.

2. The correlation between variables $j$ and $i$ is the cosine of the angle between $h_j$ and $h_i$.

3. $||h_j - h_i||^2 = s_{jj} + s_{ii} - 2s_{ji}$, that is to say, the sample variance of the difference between variables $j$ and $i$.

These three properties hold exactly for the full matrix $H$ (with all the eigenvectors), but $H_2$ only considers the first two columns of $H$. If most of the variance is explained by the two leading eigenvalues, then $H_2$ is a very good summary. Corsten and Gabriel [5] proposed the following goodness-of-fit measure for h-plotting in two dimensions, where a high measure (close to 1) indicates a better fit:

$$(\lambda_1^2 + \lambda_2^2) / \sum_j \lambda_j^2$$

Although I compute the eigenvectors of $S$, there is a clear difference with respect to Principal Component Analysis (PCA). In PCA, we are mainly interested in the representation of the $n$ objects in the component space: the scores (data matrix times eigenvectors). In fact, Euclidean distances between objects in the component space will equal their Mahalanobis distances in the observed-variable space. However, in h-plots we do not represent the $n$ objects but the $m$ variables (scaled eigenvectors are represented).

In this paper, I do not have a standard data matrix with objects and variables but a dissimilarity matrix $D$ among the $n$ objects. The value $d_{ij}$ would indicate the observed dissimilarity from the object $i$ to the object $j$. However, this dissimilarity matrix itself can be treated as a data matrix, and their variables could then be displayed as an h-plot. This new perspective allows flexibility when non-metric dissimilarities have to be represented, as we will see in the examples of the following section. I define the data matrix $X$ differently depending on whether $D$ is symmetric or not.

In the case of an asymmetric relationship being given (i.e. $d_{ij} \neq d_{ji}$), we can consider the variable measuring the dissimilarity from $j$ ($d_{j.}$) to other objects, and the variable measuring the dissimilarity from an object to $j$ ($d_{.j}$), where the dot (.) indicates an object. For each object considered in $D$, we obtain an observation of these two variables. Therefore, $X = [D'|D]$ is a $n \times 2n$ matrix (| indicates that the matrices are combined by columns). We can study the $2n$ dissimilarity variables, but only because variables are represented with the h-plot. Using PCA scores, this asymmetric relationship study is not possible.

With a symmetric dissimilarity matrix, the variable $j$ ($d_{j.} = d_{.j}$) would represent the dissimilarity with respect to $j$, which is observed for those objects in $D$ (in this case, dissimilarities from $j$ and dissimilarities to $j$ are equal, and $X = D$ is a $n \times n$ matrix). According to the previous properties of an h-plot, the Euclidean distance between $h_j$ and $h_i$ in the h-plot is approximately the sample standard deviation of the difference between vari-

ables $d_{j\cdot}$ and $d_{i\cdot}$. If these variables are similar, their difference and, therefore the standard deviation of their difference, will be small.

To my knowledge, the work most related to my approach is the use of correspondence analysis as a multidimensional scaling technique for similarity matrices [20, 27]. However, data must be in the form of similarities, so dissimilarity matrices must be first transformed to similarity matrices. Weller and Romney [27, pp. 70-76] discuss some precautions that apply when using correspondence analysis on non-frequency data. These precautions are summarized in the supplementary material, where this method is applied.

H-plot will be compared with three classical multidimensional scaling methods [10, 26]: Classical (Metric) Multidimensional Scaling (cMDS), Kruskal's Non-metric Multidimensional Scaling (isoMDS) and Sammon's Non-Linear Mapping (Sammon), and a more recent method: Isomap [24], although this method could fail when the data are spread among multiple clusters [28]. More methods are considered in the supplementary material. Note that non-metric multidimensional scaling (despite the adjective) is not designed specifically for non-metric dissimilarities. Non-metric or ordinal multidimensional scaling seeks a configuration whose distances have similar order (rank) properties. All methods take a matrix of inter-point dissimilarities as an input and create a configuration of points. A scatter plot of the points created by the methods provides a visual representation of the original dissimilarities. The Euclidean distances between those points can be computed.

Although pictures could be considered as the best method to asses the

configurations (as indicated in Borg and Groenen [2, sec. 19.7]), in order to measure and compare the configurational similarity of two configurations $X$ and $Y$, I will also calculate the congruence coefficient (a correlation coefficient about the origin), defined for symmetric dissimilarity matrices:

$$c(X,Y) = \frac{\sum_{i<j} d_{ij}(X) d_{ij}(Y)}{(\sum_{i<j} d_{ij}^2(X))^{1/2} (\sum_{i<j} d_{ij}^2(Y))^{1/2}}$$

where $d_{ij}(X)$ indicates the dissimilarity between object $i$ and $j$ in configuration $X$. The congruence coefficient (CC) ranges from 0 to 1, where 1 is achieved if $X$ and $Y$ are perfectly similar geometrically. (In geometry, two configurations are called similar if they can be brought to a complete match by rigid motions and dilations). In the experiments, I compare the original dissimilarities $X$ with the reduced configuration $Y$ obtained with each method (generally the configuration in two dimensions), using the Euclidean distance for building the interpoint distances in this configuration.

## 2.1 Comments on some benefits of the methodology

Classical multidimensional scaling methods try to preserve all pairwise proximities, whereas many of the recent nonlinear dimension reduction methods, such as Tenenbaum *et al.* [24] and Roweis and Saul [21], use only local neighborhood information to construct a global low-dimensional embedding of a hypothetical manifold near which the data fall. Both approaches could give rise to restrictive constraints in some cases when metricity is violated.

In other methods that also project the data (points in $\mathbb{R}^p$) onto a lower-dimensional manifold, such as principal surface or self-organizing map, points close together originally map close together on the manifold, but points originally far apart might also map close together [10]. However, in these methods objects must be represented as feature vectors in a vector space as they cannot work with a dissimilarity matrix. The missing vector space in my motivating problem precludes their use.

If the dissimilarity is a metric, then $|d_{ix}-d_{jx}| \leq d_{ij}$ for any object $x$ ($|\cdot|$ denotes the absolute value), due to the triangle inequality [15] ($d_{ix} \leq d_{jx}+d_{ij}$). Therefore, if $d_{ij}$ is small, the variables $d_{i\cdot}$ and $d_{j\cdot}$ will be close to each other. However, if the triangle inequality does not hold, even if $d_{ij}$ is small, variables $d_{i\cdot}$ and $d_{j\cdot}$ can be very different, and the objects $i$ and $j$ should not be represented near each other, something which is possible with h-plots. Note that in my motivating application, the triangle inequality does not hold for the dissimilarity measure. Below, I show how h-plots can visualize the intransitive dissimilarities appropriately (when the triangle inequality does not hold) in Example 1. Example 2 highlights the ease with which h-plots deal with non-Euclidean dissimilarities, comparing them with other multidimensional scaling techniques for non-metric proximities. In the supplementary material, other examples are considered. In Example 3 we will see how the representations vary when we pass from a metric to a non-metric measure, and when there are several high central objects. By a high central object, I mean an object that is similar to a large portion of other objects. Exam-

ple 4 shows how the h-plot can represent asymmetric data, and even lack of reflexivity, successfully.

**Example 1** *Let $D$ be the dissimilarity matrix, where $d_{ij}$ denotes the transit time in hours between city $i$ and city $j$, using the cheapest flight. $D$ appears in Table 1 for 4 cities: Madrid (MA) and Valencia (VL) in Spain, and Moscow (MO) and St. Petersburg (SP) in Russia. If we make a map representing these dissimilarities, we would expect to find two separate groups of neighboring cities: MA and VL, and MO and SP. MO and SP should be closer together than MA and VL, since VL's airport is not as busy as the others.*

Table 1: Dissimilarity matrix with number of hours for the cheapest flights.

| | Madrid (MA) | Valencia (VL) | Moscow (MO) | St. Petersburg (SP) |
|---|---|---|---|---|
| Madrid | 0 | 1 | 5 | 7 |
| Valencia | 1 | 0 | 10 | 12 |
| Moscow | 5 | 10 | 0 | 1.5 |
| St. Petersburg | 7 | 12 | 1.5 | 0 |

*Multidimensional methods introduced in Section 2 have been applied. Table 2 shows the Euclidean distance of the reduced configurations in one dimension and the CC. The biggest value of the CC is attained by my method, which also gives the most coherent configuration. Note that with cMDS, the distance between MA and VL (1 h. in $D$) is quite similar to that from MA to MO (5 h. in $D$); with isoMDS the distance from MA to SP (7 h. in $D$), and VL to MO (10 h. in $D$) are identical; with Sammon and Isomap (with 2 neighbors) the distance from MA to SP is even bigger than that from VL to MO, since more emphasis is put on the smaller pairwise distances.*

Table 2: Euclidean distances in the final configuration for the different methods, and their CC in one dimension.

| | cMDS | | | | isoMDS | | | | Sammon | | | | Isomap | | | | h-plot | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CC | 0.984 | | | | 0.983 | | | | 0.981 | | | | 0.974 | | | | 0.986 | | | |
| | MA | VL | MO | SP | MA | VL | MO | SP | MA | VL | MO | SP | MA | VL | MO | SP | MA | VL | MO | SP |
| MA | 0 | 4.3 | 5.8 | 7.7 | 0 | 3.2 | 6.0 | 9.2 | 0 | 1.4 | 6.4 | 8.0 | 0 | 1.0 | 5.0 | 6.5 | 0 | 2.6 | 6.3 | 7.3 |
| VL | 4.3 | 0 | 10.1 | 12.0 | 3.2 | 0 | 9.2 | 12.4 | 1.4 | 0 | 7.8 | 9.4 | 1.0 | 0 | 6.0 | 7.5 | 2.6 | 0 | 8.9 | 9.9 |
| MO | 5.8 | 10.1 | 0 | 1.9 | 6.0 | 9.2 | 0 | 3.2 | 6.4 | 7.8 | 0 | 1.6 | 5.0 | 6.5 | 0 | 1.5 | 6.3 | 8.9 | 0 | 1.1 |
| SP | 7.7 | 12.0 | 1.9 | 0 | 9.2 | 12.4 | 3.2 | 0 | 8.0 | 9.4 | 1.6 | 0 | 6.5 | 7.5 | 1.5 | 0 | 7.3 | 9.9 | 1.1 | 0 |

*I have considered only one dimension in Table 2 because with cMDS only one dimension can be obtained, as the second eigenvalue in the eigendecomposition for cMDS is negative, which indicates that D is not Euclidean [15, chapter 14], and that the distances in D cannot be reproduced exactly. A matrix is Euclidean if and only if the eigenvalues of the eigendecomposition of the classical multidimensional scaling are positive. With cMDS the relative magnitudes of those eigenvalues indicate the relative contribution of the corresponding columns in reproducing the original matrix D with the reconstructed points (in this case 9.1 -3e-14 -7e-01). Note that the negative values are not especially large in magnitude, so the configuration returned by cMDS might still reproduce D well. In Fig. 2, the configuration for two dimensions for all the methods except cMDS (with one dimension) are shown. See the supplementary material for more methods (neither of them recover the structure of the dissimilarity matrix).*

When we have a metric dissimilarity matrix and the objective is to preserve all the interpoint distances, cMDS should be the right method. The first two dimensions are the common representation, but it may be that not

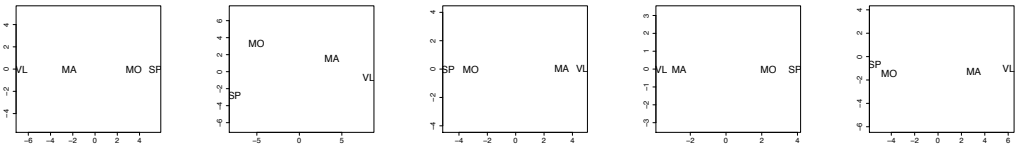


Figure 2: Scatter plot for flights: cMDS, isoMDS, Sammon, Isomap and h-plot.

all the information is in them. More dimensions could be necessary, although many dimensions are difficult to represent and understand. Hence, the two-dimensional representation may not be as good as we would wish. The other methods are alternatives according to different objectives. For example, Sammon pays more attention to the preservation of the smaller distances, which is useful for clustering. IsoMDS is more interested in preserving the rank-order of the dissimilarities. Isomap and other manifold learning techniques can be seen in a broad sense to be similar to Sammon, as it is important to retain the neighbor structure because its goal is to recover the representation of a non-linear manifold. The objective of the h-plot is not to preserve the interpoint distances exactly, or to give more weight to small distances or neighbors. Instead, h-plots aim to preserve relationships between dissimilarity variables. This point of view is especially interesting when non-metric dissimilarities are present (although also when metric dissimilarities are present, as not all the information in the dissimilarity matrix can normally be collected in two dimensions), as in this case the dissimilarities cannot be represented exactly in a Euclidean space, because the matrix is not Euclidean.

**Example 2** *In cMDS it is assumed that the dissimilarities ($d_{ij}$ in $D$) are Euclidean distances. Let $A$ be a matrix with elements $a_{ij}$ = -0.5* $d_{ij}^2$, and $B$ = $(I - n^{-1}ee')A(I - n^{-1}ee')$, where $I$ is the $n \times n$ identity matrix and $e$ is the $n \times 1$ vector with all its elements equal to unity. $D$ is Euclidean if and only if $B$ is positive semidefinite [23]. (Note that the scaled eigenvectors of $B$ are the principal coordinates in classical scaling, and note also the difference between the matrices for eigendecomposition for classical scaling and my method: $B$ = $(I - n^{-1}ee')A(I - n^{-1}ee')$ and $S = (n-1)^{-1}D'(I - n^{-1}ee')D$, respectively).*

*One approach to handling non-Euclidean pairwise data is to add a constant to the non-diagonal dissimilarities such that all eigenvalues are non-negative ([6, Ch. 2], [3]). However, the CC that I obtain in two-dimensions for Example 1 (remember that the second eigenvalue of $B$ is negative) is 0.975, less than the CC without adding a constant. The first plot in Fig. 3 shows the configuration obtained. In this figure, the distances between VL and MA, and MO and SP, are similar to the distance between MA and MO. Note that when the constant is added, the original structure and information is distorted, but in that example metric violations are not an artifact of noise; they carry relevant information. We are dealing with a genuinely non-Euclidean data set that cannot sensibly be treated as "Euclidean but noisy".*

*Recently, the problem of the information and the representation of non-Euclidean pairwise data has been studied in Laub et al [12, 13], although only symmetric dissimilarities were considered in those papers, where they paid attention to the negative part of the spectrum of $B$. They represented*

*the first two dimensions in one figure, corresponding to the first two leading eigendirections of $B$, and in another figure the last two dimensions, corresponding to the last two eigendirections of $B$, related to the metric violations. If I consider the first and last dimension in one single figure, the CC for Example 1 is 0.965, less than the other methods. In Fig. 3, the first and last two components are shown. The map that corresponds to the positive part of the eigenspectrum is a metric approximation of the dissimilarities, whereas the negative map is constructed in such a way as to correct the errors in the positive map. However, this last map is hard to interpret. According to Maaten and Hinton [14], the negative map generally contains a lot of noise. Maaten and Hinton [14] have proposed the use of multiple maps t-SNE for visualizing non-metric similarities (as usual, I have transformed the dissimilarity data into similarity data by subtracting them from a number which is larger than the largest value in the dissimilarity matrix, 12 in Example 1). They proposed this technique for representing intransitive pairwise similarities and central objects. The algorithm for computing multiple maps t-SNE is an iterative method from an initial random solution (available from http://homepage.tudelft.nl/19j49/ multiplemaps). I have run the algorithm 50 times, and the solution with the smallest error has been considered. Fig. 3 displays the two maps obtained with this method. The importance weights (represented by circles) in this example are nearly 1 (the maximum weight) for all points in the first map (the biggest weight in map 2 is 5e-05), so we can base the interpretation on map 1, where the distance between VL and*

*MA, and MO and SP, are similar to the distance between MA and MO. In both approximations specifically designed for non-metric data, we do not have one single plot, but the information should be interpreted from multiple plots, which is not always easy. With h-plot, if necessary, more dimensions can be considered in one single plot.*

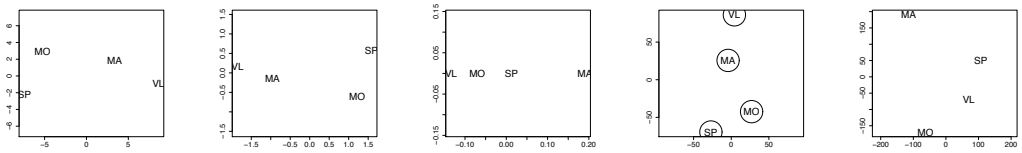


Figure 3: Scatter plot for flights: cMDS adding a constant, first two components and last two components as in Laub and Müller [12] and maps 1 and 2 as in Maaten and Hinton [14].

*I now consider the same example that Laub and Müller [12] proposed as an illustration: $D = D_1$ - $D_2$ (an 8 × 8 dissimilarity matrix). In that example, 8 objects present two salient features. They cluster into* $\{1, 2, 3, 4\}$ *and* $\{5, 6, 7, 8\}$ *according to the first feature, and into* $\{1, 3, 5, 7\}$ *and* $\{2, 4, 6, 8\}$ *according to the second.* $D_1$ *and* $D_2$ *are the dissimilarity matrices corresponding to feature 1 and 2 respectively. Instead of using two figures for the first and last components as in Laub and Müller [12], with my method all eigenvalues are positive, and I only represent the dimensions of highest variance. The configuration for my method for the first two dimensions is displayed in Fig. 4, together with the results for the method of Laub and Müller [12], multiple maps t-SNE (only map 2 is shown, as all the weights for all the points are on this map) and cMDS (Sammon and isoMDS cannot be com-*

*puted because of the negative distances, and the result is nearly zero for all the coordinates with Isomap). My method is able to discover the features in the data: no information is lost when I consider the first two dimensions, the first dimension is related to the cluster structure* $\{1, 2, 3, 4\}$ *and* $\{5, 6, 7, 8\}$*, whereas the information represented in the second dimension relates to the cluster structure* $\{1, 3, 5, 7\}$ *and* $\{2, 4, 6, 8\}$*. This structure is not recovered either with cMDS or multiple map t-SNE.*

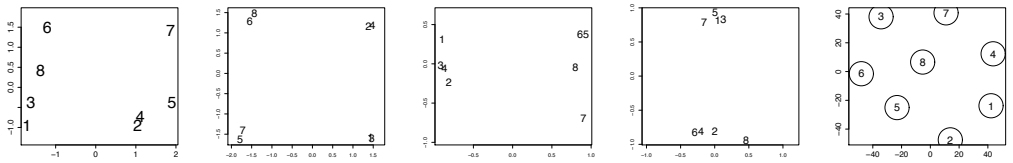

Figure 4: Scatter plot for illustration data of Laub and Müller [12]: cMDS, h-plot, first two components and last two components as in Laub and Müller [12] and map 2 as in Maaten and Hinton [14].

## 2.2 Some theoretical results of h-plots

### 2.2.1 Scaling: Effect of a linear transform of the data

Let $a_1$, $a_2$, $b_1$ and $b_2$ be scalars, and let $X$ and $Y$ be random variables, then the covariance of $a_1X + b_1$ and $a_2Y + b_2$ is given by $\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X, Y)$. Therefore, if a linear transformation $a_1X + b_1$ is applied to the dissimilarity matrix $D$, its covariance matrix is ${a_1}^2S$, where $S$ is the covariance matrix for $D$. The matrix ${a_1}^2S$ has the same eigenvectors as $S$, and the eigenvalues of $S$ are multiplied by ${a_1}^2$. As a consequence, the new $h_j$s

are the same as before the transformation but multiplied by $a_1$, not affected by $b_1$. According to property 3 of the h-plot, the squared Euclidean distance of $h_j$ and $h_i$ will be the same as before the transformation but multiplied by $a_1^2$. If the scale of the dissimilarities is linearly modified, the resulting configuration does not change in the sense that the visual configuration will be the same as before, and only the scale of the axes is changed (multiplied by $a_1$). In practice, it does not matter if the dissimilarity is expressed in hours or minutes, or in kilometers or meters.

### 2.2.2 Noise sensitivity

It is well-known that extreme observations, outliers, may have a considerable influence on the covariance matrix structure, and therefore they can influence the h-plot. To counteract this influence, robust h-plots can be built with M-estimates, as explained in Daigle and Rivest [7].

# 3 Corneal endothelia application

I have applied the method presented to my motivating problem. The corneal endothelia analyzed belong to 153 individuals of between 17 and 84 years old. Here, point patterns are the objects that form the data set. Ten different dissimilarities were contemplated with univariate point patterns by Ayala *et al.* [1]. I have considered dissimilarities based on the log-rank statistic applied to the nearest-neighbor distances, the same dissimilarities with triple

points used in the clinical application in Ayala *et al.* [1], where the best results in clustering were obtained with that dissimilarity. This dissimilarity measure is symmetric, but it is not a metric since the triangle inequality does not hold. I have also considered this dissimilarity for the three simulated experiments in Ayala *et al.* [1]. Details of the dissimilarities and the experiments with their results are given in the supplementary material.

Figure 5 shows the results for the established methods and my method, both using the original dissimilarities and their ranks (the original dissimilarities are replaced with their sample ranks). If we have in mind cluster and pattern detection, then an expansion or contraction of the configuration could be more useful [23]. For this reason, I also consider the ranking of the dissimilarities instead of the original dissimilarity values. Analogously, other transformations could be considered such as raising dissimilarities to a power [18]. The unhealthy cases obtained in Ayala *et al.* [1] are represented by red triangles, while black circles are healthy cases. I have used $\epsilon = 1$ with the Isomap algorithm to obtain an appropriate representation. Note that seven points have been automatically removed (and they are not displayed) with this algorithm, for being considered as outliers. The CC for Isomap has also been computed without these points. Table 3 shows the CCs. Even though Isomap does not take into account some extreme points, its coefficient is the smallest. However, all coefficients are quite high. The goodness-of-fit for my method is 81.59% and 99.85% for one and two dimensions, respectively. We can differentiate both groups in the figures, although the boundary between

them is not clear. This is quite reasonable since the status could be considered as an ordered factor, from the most severe unhealthy cases to the most healthy cases. Note that pictures (a), (b) and (e) in Fig. 5 seem similar, but the healthy cases in the h-plot are not only discriminated by the first dimension but also by the second dimension. I have carried out several ANOVA-style analyses using the functions in Oksanen *et al.* [16]: *adonis* and *anosim*. The Euclidean distances in the reduced configuration for each method are explained by the factor status (healthy or unhealthy). For all methods and functions, the effect is significant with p-value = 0.001. However, the $R$ statistic of *anosim* (as $R$ approaches 1, there is more dissimilarity between groups) and $R - squared$ for *adonis* are different for each method. Their values are in Table 3. The biggest values for both are obtained with the h-plot. With the h-plot of the dissimilarity ranks, the values are even bigger: 0.915 and 0.733 for *anosim* and *adonis* respectively.

Table 3: CCs, *anosim* statistic $R$ and $R^2$ of *adonis* with endothelia.

| | cMDS | isoMDS | Sammon | Isomap | h-plot |
|---|---|---|---|---|---|
| CC | 0.935 | 0.929 | 0.894 | 0.881 | 0.922 |
| $R$ *anosim* | 0.578 | 0.761 | 0.708 | 0.674 | 0.829 |
| $R^2$ *adonis* | 0.457 | 0.614 | 0.619 | 0.541 | 0.704 |

# 4 Conclusions

Despite non-Euclidean or non-metric measures becoming more popular, there are not many methods in literature for the specific representation (without

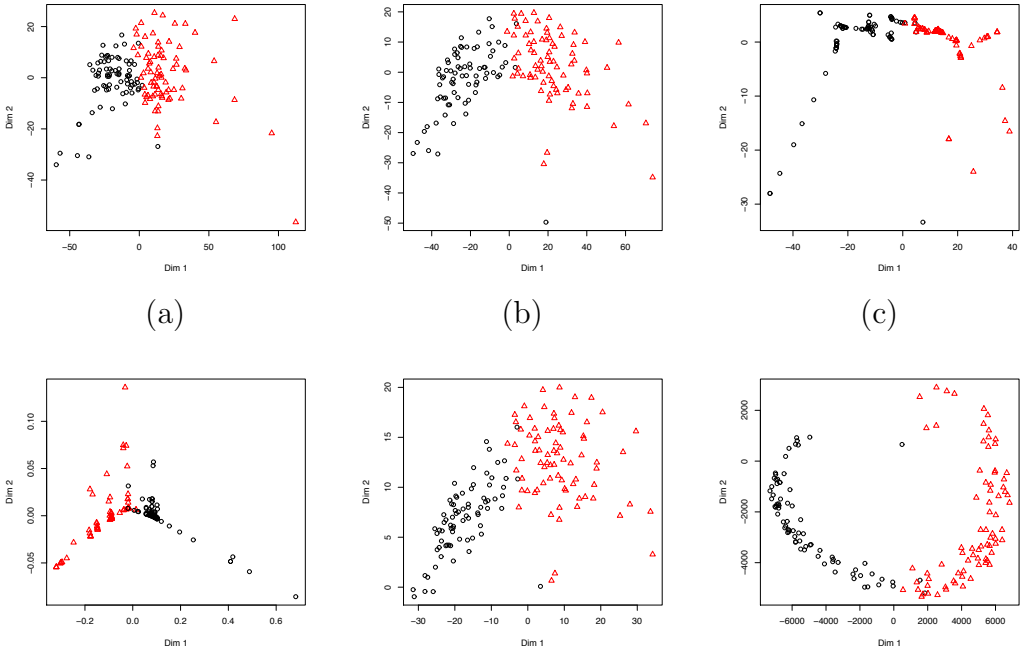


Figure 5: Endothelia: (a) cMDS, (b) isoMDS, (c) Sammon, (d) Isomap, and h-plot using: (e) the original dissimilarities, and (f) the dissimilarity ranks.

data transformation) of non-Euclidean pairwise data. When data are inherently non-metric, we should not enforce metricity, as real information could be lost. I have presented a method for displaying (non-metric) dissimilarity matrices, based on h-plots. It handles asymmetric data, and even lack of reflexivity, naturally. Its good performance is shown through several examples, and particularly in my motivating application: the analysis of human corneal endothelia, where the dissimilarity was not a metric. Furthermore, this method is very simple to implement and computationally efficient. The representation goodness can also be easily assessed.

With the h-plot, we represent second order differences between variables that indicate dissimilarity with respect to an object. In future work, higher order differences or looking for associations [19] could be considered, although the simplicity of the present method could be lost.

# Acknowledgment

This work has been partially supported by Grants CICYT TIN2009-14392-C02-01 and MTM2009-14500-C02-02, GV/2011/004 and Bancaixa-UJI P11A2009-02. The author would like to thank D. Peña and the reviewers for their very constructive suggestions, and M. A. García and J. Epifanio for their support.

# References

[1] Ayala, G., Epifanio, I., Simó, A., and Zapater, V. (2006). Clustering of spatial point patterns. *Computational Statistics & Data Analysis*, **50**(4), 1016–1032.

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

[3] Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, **48**, 343–349.

[4] Chen, L. and Buja, A. (2009). Local multidimensional scaling for nonlin-

ear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association*, **104**(485), 444–472.

[5] Corsten, L. C. A. and Gabriel, K. R. (1976). Graphical exploration in comparing variance matrices. *Biometrics*, **32**(4), 851–863.

[6] Cox, T. F. and Cox, M. A. A. (1994). *Multidimensional Scaling*. Chapman and Hall.

[7] Daigle, G. and Rivest, L. P. (1992). A robust biplot. *The Canadian Journal of Statistics*, **20**(3), 241–255.

[8] Gabriel, K. R. (1971). The biplot-graphic display of matrices with applications to principal component analysis. *Biometrika*, **58**, 453–467.

[9] Gower, J. C. and Hand, D. J. (1996). *Biplots*. Chapman and Hall.

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

[11] Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques. Regression, Classification, and Manifold Learning*. Springer-Verlag.

[12] Laub, J. and Müller, K. R. (2004). Feature discovery in non-metric pairwise data. *Journal of Machine Learning Research*, **5**, 801–818.

[13] Laub, J., Roth, V., Buhmann, J. M., and Müller, K. R. (2006). On the information and representation of non-Euclidean pairwise data. *Pattern Recognition*, **39**(10), 1815–1826.

[14] Maaten, L. and Hinton, G. (2012). Visualizing non-metric similarities in multiple maps. *Machine Learning*, **87**(1), 33–55.

[15] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1989). *Multivariate Analysis*. Academic Press.

[16] Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2012). *vegan: Community Ecology Package*. R package version 2.0-4.

[17] Pękalska, E. and Duin, R. P. W. (2005). *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications*. World Scientific.

[18] Podani, J. and Miklós, I. (2002). Resemblance coefficients and the horseshoe effect in principal coordinates analysis. *Ecology*, **83**(12), 3331–3343.

[19] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science*, **334**(6062), 1518–1524.

[20] Romney, A. K., Moore, C. C., and Brazill, T. J. (1998). *Visualization of Categorical Data*, chapter Correspondence Analysis as a multidimesnional scaling technique for nonfrequency similarity matrices, pages 329–345. Academic Press, London.

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

[22] Schölkopf, B., Smola, A., and Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, **10**(5), 1299–1319.

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

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

[25] Trosset, M. W. and Priebe, C. E. (2008). The out-of-sample problem for classical multidimensional scaling. *Computational Statistics & Data Analysis*, **52**(10), 4635–4642.

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

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

[28] 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.