Metric Multidimensional Scaling (MDS): Analyzing Distance Matrices

Hervé Abdi[1]

1 Overview

Metric multidimensional scaling (MDS) transforms a distance matrix into a set of coordinates such that the (Euclidean) distances derived from these coordinates approximate as well as possible the original distances. The basic idea of MDS is to transform the distance matrix into a cross-product matrix and then to find its eigen-decomposition which gives a principal component analysis (PCA). Like PCA, MDS can be used with supplementary or illustrative elements which are projected onto the dimensions after they have been computed.

2 An example

The example is derived from O'Toole, Jiang, Abdi, and Haxby (2005), in which the authors used a combination of principal component

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: `herve@utdallas.edu` `http://www.utd.edu/~herve`
Revised May 26, 2009. (Equations 17, 18, and 21)

analysis and neural networks to analyze brain imaging data. In this study 6 subjects were scanned using *f*MRI when they were watching pictures from 8 categories (faces, houses, cats, chairs, shoes, scissors, bottles and scrambled images). The authors computed for each subject a distance matrix corresponding to how well they could predict the type of pictures that this subject was watching from his/her brain scans. The distance used was d' (see entry) which expresses the discriminability between categories.

O'Toole *et al.,* give two distance matrices. The first one is the average distance matrix computed from the brain scans of all 6 subjects. The authors also give a distance matrix derived directly from the pictures watched by the subjects. The authors computed this distance matrix with the same algorithm that they used for the brain scans, they just substituted images to brain scans.

We will use these two matrices to review the basic of multidimensional scaling: namely how to transform a distance matrix into a cross-product matrix and how to project a set of supplementary observations onto the space obtained by the original analysis.

3 Multidimensional Scaling: Eigen-analysis of a distance matrix

PCA is obtained by performing the eigen-decomposition of a matrix. This matrix can be a correlation matrix (*i.e.,* the variables to be analyzed are centered and normalized), a covariance matrix (*i.e.,* the variables are centered but not normalized), or a cross-product matrix (*i.e.,* the variables are neither centered nor normalized). A distance matrix cannot be analyzed directly using the eigen-decomposition (because distance matrices are not positive semi-definite matrices), but it can be transformed into an equivalent cross-product matrix which can then be analyzed.

3.1 Transforming a distance matrix into a cross-product matrix

In order to transform a distance matrix into a cross-product matrix, we start from the observation that the scalar product between two vectors can easily be transformed into a distance (the scalar product between vectors corresponds to a cross-product matrix). Let us start with some definitions. Suppose that **a** and **b** are two vectors with I elements, the Euclidean distance between these two vectors is computed as

$$d^2(\mathbf{a},\mathbf{b}) = (\mathbf{a}-\mathbf{b})^\mathsf{T}(\mathbf{a}-\mathbf{b}) \; . \tag{1}$$

This distance can be rewritten in order to isolate the scalar product between vectors **a** and **b**:

$$d^2(\mathbf{a},\mathbf{b}) = (\mathbf{a}-\mathbf{b})^\mathsf{T}(\mathbf{a}-\mathbf{b}) = \mathbf{a}^\mathsf{T}\mathbf{a}+\mathbf{b}^\mathsf{T}\mathbf{b}-2\times(\mathbf{a}^\mathsf{T}\mathbf{b}) \; , \tag{2}$$

where $\mathbf{a}^\mathsf{T}\mathbf{b}$ is the scalar product between **a** and **b**.

If the data are stored into an I by J data matrix denoted **X** (where I observations are described by J variables), the between observations cross product matrix is then obtained as

$$\underset{I\times I}{\mathbf{S}} = \underset{I\times J}{\mathbf{X}} \times \underset{J\times I}{\mathbf{X}}^\mathsf{T} \; . \tag{3}$$

A distance matrix can be computed directly from the cross-product matrix as

$$\underset{I\times I}{\mathbf{D}} = \underset{I\times 1}{\mathbf{s}}\,\underset{1\times I}{\mathbf{1}}^\mathsf{T} + \underset{I\times 1}{\mathbf{1}}\,\underset{1\times I}{\mathbf{s}}^\mathsf{T} - 2\,\underset{I\times I}{\mathbf{S}} \; . \tag{4}$$

(Note that the elements of **D** gives the *squared* Euclidean distance between rows of **S**)

This equation shows that an Euclidean distance matrix can be computed from a cross-product matrix. In order to perform MDS on a set of data, the main idea is to “revert" Equation 4 in order to obtain a cross-product matrix from a distance matrix. There is one problem when implementing this idea, namely that *different* cross-product matrices can give the *same* distance. This can happen because distances are invariant for any change of origin. Therefore, in order to revert the equation we need to impose an origin for the computation of the distance. An obvious choice is to choose the origin of the distance as the center of gravity of the dimensions. With this constraint, the cross-product matrix is obtained as follows.

First define a mass vector denoted $\mathbf{m}$ whose I elements give the mass of the I rows of matrix $\mathbf{D}$. These elements are all positive and their sum is equal to one:

$$\underset{1\times}{\mathbf{m}^{\top}}\underset{I\times 1}{\mathbf{1}} = 1 \,. \tag{5}$$

When all the rows have equal importance, each element is equal to $\frac{1}{I}$.

Second, define an $I \times I$ centering matrix denoted $\boldsymbol{\Xi}$ (read "big Xi") equal to

$$\underset{I\times I}{\boldsymbol{\Xi}} = \underset{I\times I}{\mathbf{I}} - \underset{I\times 1}{\mathbf{1}}\underset{1\times I}{\mathbf{m}^{\top}} \,. \tag{6}$$

Finally, the cross-product matrix is obtained from matrix $\mathbf{D}$ as:

$$\underset{I\times I}{\mathbf{S}} = -\tfrac{1}{2}\boldsymbol{\Xi}\mathbf{D}\boldsymbol{\Xi}^{\top} \,. \tag{7}$$

The eigen-decomposition of this matrix gives

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\top} \tag{8}$$

with

$$\mathbf{U}^{\top}\mathbf{U} = \mathbf{I} \text{ and } \boldsymbol{\Lambda} \text{ diagonal matrix of eigenvalues.} \tag{9}$$

(see appendix for a proof).

The scores (*i.e.,* the projection of the rows on the principal components of the analysis of $\mathbf{S}$) are obtained as

$$\mathbf{F} = \mathbf{M}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \text{ (with } \mathbf{M} = \operatorname{diag}\{\mathbf{m}\}) \tag{10}$$

The scores have the properties that their variance is equal to the eigenvalues:

$$\mathbf{F}^{\top}\mathbf{M}\mathbf{F} = \boldsymbol{\Lambda} \,. \tag{11}$$

Table 1: The d' matrix from O'Toole *et al.*, (2005). This matrix gives the d' obtained for the discrimination between categories based upon the brain scans. These data are obtained by averaging 12 data tables (2 per subject).

	Face	House	Cat	Chair	Shoe	Scissors	Bottle	Scrambled
Face	0.00	3.47	1.79	3.00	2.67	2.58	2.22	3.08
House	3.47	0.00	3.39	2.18	2.86	2.69	2.89	2.62
Cat	1.79	3.39	0.00	2.18	2.34	2.09	2.31	2.88
Chair	3.00	2.18	2.18	0.00	1.73	1.55	1.23	2.07
Shoes	2.67	2.86	2.34	1.73	0.00	1.44	1.29	2.38
Scissors	2.58	2.69	2.09	1.55	1.44	0.00	1.19	2.15
Bottle	2.22	2.89	2.31	1.23	1.29	1.19	0.00	2.07
Scrambled	3.08	2.62	2.88	2.07	2.38	2.15	2.07	0.00

Table 2: The d' matrix from O'Toole *et al.*, (2005). This matrix gives the d' obtained for the discrimination between categories based upon the images watched by the subjects.

	Face	House	Cat	Chair	Shoe	Scissors	Bottle	Scrambled
Face	0.00	4.52	4.08	4.08	4.52	3.97	3.87	3.73
House	4.52	0.00	2.85	4.52	4.52	4.52	4.08	4.52
Cat	4.08	2.85	0.00	1.61	2.92	2.81	1.96	3.17
Chair	4.08	4.52	1.61	0.00	2.82	2.89	2.91	3.97
Shoe	4.52	4.52	2.92	2.82	0.00	3.55	3.26	4.52
Scissors	3.97	4.52	2.81	2.89	3.55	0.00	2.09	3.26
Bottle	3.87	4.08	1.96	2.91	3.26	2.09	0.00	1.50
Scrambled	3.73	4.52	3.17	3.97	4.52	3.26	1.50	0.00

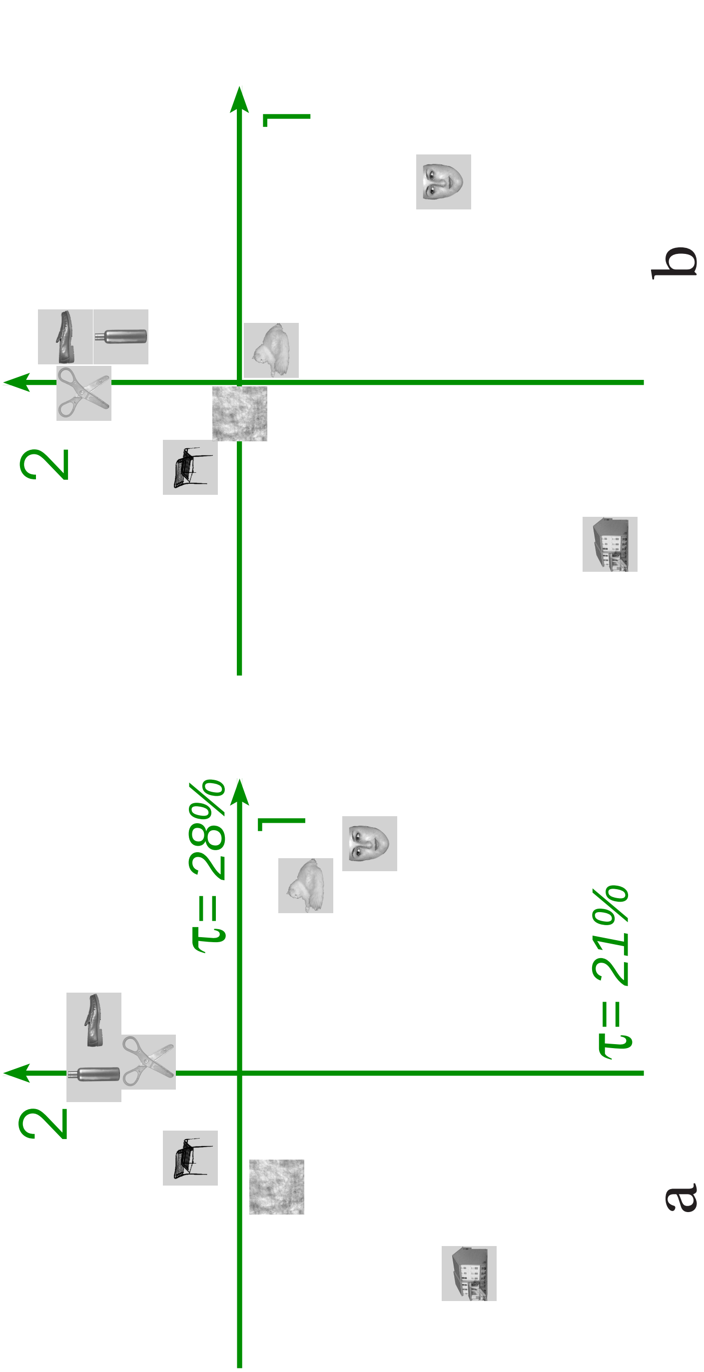

Figure 1: (a) Multidimensional scaling of the subjects's distance table. (b) Projection of the image distance table as supplementary elements in the subjects' space (Distance from Tables 1 and 2).

3.2 Example

To illustrate the transformation of the distance matrix, we will use the distance matrix derived from the brain scans given in Table 1:

$$\mathbf{D} = \begin{bmatrix} 0.00 & 3.47 & 1.79 & 3.00 & 2.67 & 2.58 & 2.22 & 3.08 \\ 3.47 & 0.00 & 3.39 & 2.18 & 2.86 & 2.69 & 2.89 & 2.62 \\ 1.79 & 3.39 & 0.00 & 2.18 & 2.34 & 2.09 & 2.31 & 2.88 \\ 3.00 & 2.18 & 2.18 & 0.00 & 1.73 & 1.55 & 1.23 & 2.07 \\ 2.67 & 2.86 & 2.34 & 1.73 & 0.00 & 1.44 & 1.29 & 2.38 \\ 2.58 & 2.69 & 2.09 & 1.55 & 1.44 & 0.00 & 1.19 & 2.15 \\ 2.22 & 2.89 & 2.31 & 1.23 & 1.29 & 1.19 & 0.00 & 2.07 \\ 3.08 & 2.62 & 2.88 & 2.07 & 2.38 & 2.15 & 2.07 & 0.00 \end{bmatrix} . \tag{12}$$

The elements of the mass vector $\mathbf{m}$ are all equal to $\frac{1}{8}$;

$$\mathbf{m}^\mathsf{T} = \begin{bmatrix} .125 & .125 & .125 & .125 & .125 & .125 & .125 & .125 \end{bmatrix} . \tag{13}$$

The centering matrix is equal to:

$$\underset{8\times 8}{\boldsymbol{\Xi}} = \begin{bmatrix} .875 & -.125 & -.125 & -.125 & -.125 & -.125 & -.125 & -.125 \\ -.125 & .875 & -.125 & -.125 & -.125 & -.125 & -.125 & -.125 \\ -.125 & -.125 & .875 & -.125 & -.125 & -.125 & -.125 & -.125 \\ -.125 & -.125 & -.125 & .875 & -.125 & -.125 & -.125 & -.125 \\ -.125 & -.125 & -.125 & -.125 & .875 & -.125 & -.125 & -.125 \\ -.125 & -.125 & -.125 & -.125 & -.125 & .875 & -.125 & -.125 \\ -.125 & -.125 & -.125 & -.125 & -.125 & -.125 & .875 & -.125 \\ -.125 & -.125 & -.125 & -.125 & -.125 & -.125 & -.125 & .875 \end{bmatrix} .$$

The cross product matrix is then equal to

$$\mathbf{S} = \begin{bmatrix} 1.34 & -0.31 & 0.34 & -0.46 & -0.25 & -0.26 & -0.12 & -0.29 \\ -0.31 & 1.51 & -0.38 & 0.03 & -0.26 & -0.24 & -0.37 & 0.02 \\ 0.34 & -0.38 & 1.12 & -0.16 & -0.19 & -0.14 & -0.27 & -0.31 \\ -0.46 & 0.03 & -0.16 & 0.74 & -0.08 & -0.05 & 0.07 & -0.09 \\ -0.25 & -0.26 & -0.19 & -0.08 & 0.83 & 0.05 & 0.09 & -0.20 \\ -0.26 & -0.24 & -0.14 & -0.05 & 0.05 & 0.71 & 0.08 & -0.15 \\ -0.12 & -0.37 & -0.27 & 0.07 & 0.09 & 0.08 & 0.65 & -0.13 \\ -0.29 & 0.02 & -0.31 & -0.09 & -0.20 & -0.15 & -0.13 & 1.15 \end{bmatrix} .$$

The eigen-decomposition of **S** gives

$$\mathbf{U} = \begin{bmatrix} 0.60 & -0.36 & -0.10 & 0.48 & -0.23 & 0.02 & 0.30 \\ -0.52 & -0.64 & 0.36 & 0.14 & 0.10 & -0.06 & -0.18 \\ 0.48 & -0.17 & 0.10 & -0.67 & 0.24 & 0.04 & -0.30 \\ -0.23 & 0.16 & 0.20 & -0.38 & -0.54 & 0.29 & 0.49 \\ -0.02 & 0.39 & 0.19 & 0.28 & 0.61 & 0.47 & 0.14 \\ -0.03 & 0.32 & 0.11 & -0.00 & 0.14 & -0.83 & 0.23 \\ 0.00 & 0.38 & 0.02 & 0.25 & -0.43 & 0.04 & -0.69 \\ -0.28 & -0.08 & -0.87 & -0.09 & 0.11 & 0.04 & 0.02 \end{bmatrix} \tag{14}$$

and

$$\mathbf{\Lambda} = \begin{bmatrix} 2.22 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.72 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.23 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.79 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.69 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.39 \end{bmatrix} . \tag{15}$$

As in PCA, the eigenvalues are often transformed into percentage of explained variance (or inertia) in order to make their interpretation easier. Here, for example, we find that the first dimension "explains" 28% of the variance of the distances (*i.e.*, $\frac{2.22}{2.22+\cdots+0.39} = .28$).

We obtain the following matrix of scores.

$$\mathbf{F} = \begin{bmatrix} 2.53 & -1.35 & -0.30 & 1.36 & -0.58 & 0.04 & 0.53 \\ -2.19 & -2.37 & 1.13 & 0.39 & 0.24 & -0.15 & -0.32 \\ 2.04 & -0.63 & 0.32 & -1.90 & 0.61 & 0.10 & -0.52 \\ -0.97 & 0.61 & 0.62 & -1.09 & -1.35 & 0.68 & 0.86 \\ -0.10 & 1.44 & 0.59 & 0.81 & 1.53 & 1.10 & 0.25 \\ -0.13 & 1.18 & 0.33 & -0.00 & 0.35 & -1.96 & 0.40 \\ 0.02 & 1.41 & 0.05 & 0.70 & -1.09 & 0.09 & -1.22 \\ -1.20 & -0.29 & -2.74 & -0.27 & 0.28 & 0.10 & 0.03 \end{bmatrix} .$$

Figure 1*a* displays the projection of the categories on the first two dimensions. The first dimension explains 28% of the variance of the distance, it can be interpreted as the opposition of the face and cat categories to the house category (these categories are the ones most easily discriminated in the scans). The second dimension, which explains 21% of the variance, separates the small objects from the other categories.

3.3 Multidimensional scaling: Supplementary elements

After we have computed the MDS solution, it is possible to project supplementary or illustrative elements onto this solution. To illustrate this procedure, we will project the distance matrix obtained from the pictures (see Table 2) onto the space defined by the analysis of the brain scans.

The number of supplementary elements is denoted by I_{sup}. For each supplementary elements, we need the values of its distances to all the I active elements. We can store these distances in an $I \times I_{\text{sup}}$ supplementary distance matrix denoted $\mathbf{D}_{\text{sup}}$. So, for our example, we have:

$$\mathbf{D}_{\text{sup}} = \begin{bmatrix} 0.00 & 4.52 & 4.08 & 4.08 & 4.52 & 3.97 & 3.87 & 3.73 \\ 4.52 & 0.00 & 2.85 & 4.52 & 4.52 & 4.52 & 4.08 & 4.52 \\ 4.08 & 2.85 & 0.00 & 1.61 & 2.92 & 2.81 & 1.96 & 3.17 \\ 4.08 & 4.52 & 1.61 & 0.00 & 2.82 & 2.89 & 2.91 & 3.97 \\ 4.52 & 4.52 & 2.92 & 2.82 & 0.00 & 3.55 & 3.26 & 4.52 \\ 3.97 & 4.52 & 2.81 & 2.89 & 3.55 & 0.00 & 2.09 & 3.26 \\ 3.87 & 4.08 & 1.96 & 2.91 & 3.26 & 2.09 & 0.00 & 1.50 \\ 3.73 & 4.52 & 3.17 & 3.97 & 4.52 & 3.26 & 1.50 & 0.00 \end{bmatrix} . \quad (16)$$

The first step is to transform $\mathbf{D}_{\text{sup}}$ into a cross-product matrix denoted $\mathbf{S}_{\text{sup}}$. This is done by first computing the difference for each supplementary column and the average distance vector and then centering the rows with the same centering matrix that was used previously to transform the distance of the active elements. Specifically, the supplementary cross-product matrix is obtained as:

$$\mathbf{S}_{\text{sup}} = -\frac{1}{2}\boldsymbol{\Xi}\left(\mathbf{D}_{\text{sup}} - \mathbf{Dm1}^{\mathsf{T}}\right) \quad (17)$$

(where $\mathbf{1}$ is an I_{sup} by 1 vector of ones; note, also, that when $\mathbf{D}_{\text{sup}} =$

D, Equation 17 reduces to Equation 7). For our example, this gives:

$$\mathbf{S}_{\text{sup}} = \begin{bmatrix} 1.97 & -0.24 & -0.66 & -0.45 & -0.46 & -0.37 & -0.54 & -0.15 \\ -0.21 & 2.10 & 0.04 & -0.58 & -0.38 & -0.57 & -0.56 & -0.47 \\ -0.19 & 0.48 & 1.27 & 0.68 & 0.23 & 0.09 & 0.31 & 0.01 \\ -0.38 & -0.55 & 0.27 & 1.29 & 0.09 & -0.13 & -0.36 & -0.58 \\ -0.55 & -0.50 & -0.33 & -0.07 & 1.55 & -0.42 & -0.49 & -0.80 \\ -0.33 & -0.56 & -0.34 & -0.17 & -0.29 & 1.29 & 0.03 & -0.24 \\ -0.32 & -0.37 & 0.05 & -0.21 & -0.18 & 0.22 & 1.05 & 0.61 \\ 0.01 & -0.34 & -0.30 & -0.49 & -0.55 & -0.12 & 0.55 & 1.61 \end{bmatrix} . \tag{18}$$

The next step is to project the matrix $\mathbf{S}_{\text{sup}}$ onto the space defined by the analysis of the active distance matrix. We denote by $\mathbf{F}_{\text{sup}}$ the matrix of projection of the supplementary elements. Its computational formula is obtained by first combining Equations 10 and 8 in order to get

$$\mathbf{F} = \mathbf{S}^{\mathsf{T}}\mathbf{M}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{-\frac{1}{2}} , \tag{19}$$

and then substituting $\mathbf{S}_{\text{sup}}$ for $\mathbf{S}$ and simplifying. This gives

$$\mathbf{F}_{\text{sup}} = \mathbf{S}_{\text{sup}}^{\mathsf{T}}\mathbf{M}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \mathbf{S}_{\text{sup}}^{\mathsf{T}}\mathbf{F}\boldsymbol{\Lambda}^{-1} . \tag{20}$$

For our example, this equation gives the following values:

$$\mathbf{F}_{\text{sup}} = \begin{bmatrix} 2.48 & -2.27 & -1.29 & 2.69 & -1.78 & -0.19 & 2.57 \\ -1.44 & -4.13 & 2.19 & -0.40 & 1.30 & -0.23 & -3.66 \\ 0.46 & -0.33 & 1.08 & -3.73 & 0.04 & 0.80 & -2.79 \\ 0.40 & 1.10 & 1.39 & -3.59 & -1.63 & 1.72 & 2.23 \\ 0.27 & 1.86 & 1.76 & -0.03 & 3.17 & 3.35 & 0.80 \\ 0.29 & 1.73 & -0.06 & -0.91 & -0.17 & -4.32 & -0.09 \\ 0.11 & 1.34 & -1.90 & -0.95 & -1.10 & -0.88 & -5.03 \\ -0.27 & -0.05 & -4.71 & -0.43 & -0.97 & -0.77 & -3.70 \end{bmatrix} . \tag{21}$$

Figure 1*b* displays the projection of the supplementary categories on the first two dimensions. Comparing plots *a* and *b* shows that an analysis of the pictures reveals a general map very similar to the analysis of the brain scans with only one major difference: The cat category for the images moves to the center of the space. This suggests that the cat category is interpreted by the subjects as being face-like (*i.e.,* “cats have faces").

4 Analyzing non-metric data

Metric MDS is adequate only when dealing with distances (see Togerson, 1958). In order to accommodate weaker measurements (called dissimilarities) non-metric MDS is adequate. It derives an Euclidean distance approximation using only the ordinal information from the data (Shepard, 1966; for a recent thorough review, see Borg & Groenen, 1997).

Appendix: Proof

We start with an $I \times I$ distance matrix $\mathbf{D}$, and an $I \times 1$ vector of mass (whose elements are all positive or zero and whose sum is equal to 1) denoted $\mathbf{m}$ and such that

$$\underset{1\times}{\mathbf{m}^{\mathsf{T}}} \underset{I\times 1}{\mathbf{1}} = 1 \,. \tag{22}$$

The centering matrix is equal to

$$\underset{I\times I}{\boldsymbol{\Xi}} = \underset{I\times I}{\mathbf{I}} - \underset{I\times 1}{\mathbf{1}} \underset{\times I}{\mathbf{m}^{\mathsf{T}}} \,. \tag{23}$$

We want to show that the following cross-product matrix

$$\underset{I\times I}{\mathbf{S}} = -\tfrac{1}{2}\boldsymbol{\Xi}\mathbf{D}\boldsymbol{\Xi}^{\mathsf{T}} \,, \tag{24}$$

will give back the original distance matrix when the distance matrix is computed as:

$$\underset{I\times I}{\mathbf{D}} = \underset{I\times 1}{\mathbf{s}} \underset{\times I}{\mathbf{1}^{\mathsf{T}}} + \underset{I\times 1}{\mathbf{1}} \underset{\times I}{\mathbf{s}^{\mathsf{T}}} - 2 \underset{I\times I}{\mathbf{S}} \,. \tag{25}$$

In order to do so, we need to choose an origin for the coordinates (because several coordinates systems will give the same distance matrix). A natural choice is to assume that the data are centered (*i.e.,* the mean of each original variable is equal to zero). There we assume that the mean vector, denoted $\mathbf{c}$ computed as:

$$\underset{J\times 1}{\mathbf{c}} = \underset{J\times}{\mathbf{X}^{\mathsf{T}}} \underset{I\times 1}{\mathbf{m}} \,, \tag{26}$$

(for some data matrix **X**). Because the origin of the space is located at the center of gravity, its coordinates are equal to $\mathbf{c} = \mathbf{0}$. The cross-product matric can therefore be computed as

$$\underset{I\times I}{\mathbf{S}} = \left(\underset{I\times J}{\mathbf{X}} - \underset{I\times 1}{\mathbf{1}}\,\underset{1\times J}{\mathbf{c}}^{\mathsf{T}}\right)\left(\underset{I\times J}{\mathbf{X}} - \underset{I\times 1}{\mathbf{1}}\,\underset{1\times J}{\mathbf{c}}^{\mathsf{T}}\right)^{\mathsf{T}}$$

$$= \left(\underset{I\times J}{\mathbf{X}} - \underset{I\times 1}{\mathbf{1}}\,\underset{1\times J}{\mathbf{c}}^{\mathsf{T}}\right)\left(\underset{J\times I}{\mathbf{X}^{\mathsf{T}}} - \underset{J\times 1}{\mathbf{c}}\,\underset{1\times I}{\mathbf{1}}^{\mathsf{T}}\right). \tag{27}$$

First, we assume that there exists a matrix denoted **S** such that Equation 25) is satisfied. Then we plug Equation 25 into Equation 24, develop and simplify in order to get

$$-\tfrac{1}{2}\boldsymbol{\Xi}\mathbf{D}\boldsymbol{\Xi}^{\mathsf{T}} = -\tfrac{1}{2}\boldsymbol{\Xi}\mathbf{s}\mathbf{1}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}} - \tfrac{1}{2}\boldsymbol{\Xi}\mathbf{1}\mathbf{s}^{\mathsf{T}}\boldsymbol{\Xi}^{\mathsf{T}} + \boldsymbol{\Xi}\mathbf{S}\boldsymbol{\Xi}^{\mathsf{T}}. \tag{28}$$

Then we show that the terms $\boldsymbol{\Xi}\left(\mathbf{s}\mathbf{1}^{\mathsf{T}}\right)\boldsymbol{\Xi}^{\mathsf{T}}$ and $\boldsymbol{\Xi}\left(\mathbf{1}\mathbf{s}^{\mathsf{T}}\right)\boldsymbol{\Xi}^{\mathsf{T}}$ are null because::

$$\begin{aligned}
\left(\mathbf{s}\mathbf{1}^{\mathsf{T}}\right)\boldsymbol{\Xi}^{\mathsf{T}} &= \mathbf{s}\mathbf{1}^{\mathsf{T}}\left(\mathbf{I} - \mathbf{1}\mathbf{m}^{\mathsf{T}}\right)^{\mathsf{T}} \\
&= \mathbf{s}\mathbf{1}^{\mathsf{T}}\left(\mathbf{I} - \mathbf{m}\mathbf{1}^{\mathsf{T}}\right) \\
&= \mathbf{s}\mathbf{1}^{\mathsf{T}} - \mathbf{s}\mathbf{1}^{\mathsf{T}}\mathbf{m}\mathbf{1}^{\mathsf{T}} \quad \text{(but from Equation 22: } \mathbf{1}^{\mathsf{T}}\mathbf{m} = 1\text{)} \\
&= \mathbf{s}\mathbf{1}^{\mathsf{T}} - \mathbf{s}\mathbf{1}^{\mathsf{T}} \\
&= \underset{I\times I}{\mathbf{0}}.
\end{aligned} \tag{29}$$

The last thing to show now is that the term $\boldsymbol{\Xi}\mathbf{S}\boldsymbol{\Xi}^{\mathsf{T}}$ is equal to **S**. This is shown by developing:

$$\begin{aligned}
\boldsymbol{\Xi}\mathbf{S}\boldsymbol{\Xi}^{\mathsf{T}} &= \left(\mathbf{I} - \mathbf{1}\mathbf{m}^{\mathsf{T}}\right)\mathbf{S}\left(\mathbf{I} - \mathbf{m}\mathbf{1}^{\mathsf{T}}\right) \\
&= \mathbf{S} - \mathbf{S}\mathbf{m}\mathbf{1}^{\mathsf{T}} - \mathbf{1}\mathbf{m}^{\mathsf{T}}\mathbf{S} + \mathbf{1}\mathbf{m}^{\mathsf{T}}\mathbf{S}\mathbf{m}\mathbf{1}^{\mathsf{T}}.
\end{aligned} \tag{30}$$

Because

$$\left(\mathbf{X}^{\mathsf{T}} - \mathbf{c}\mathbf{1}^{\mathsf{T}}\right)\mathbf{m} = \mathbf{X}^{\mathsf{T}}\mathbf{m} - \mathbf{c}\mathbf{1}^{\mathsf{T}}\mathbf{m} \quad (\textit{cf.}\text{ Equations 26 and 22})$$

$$= \mathbf{c} - \mathbf{c}$$

$$= \underset{I \times 1}{\mathbf{0}}, \tag{31}$$

we get (*cf.* Equation 27):

$$\mathbf{Sm} = \left(\mathbf{X} - \mathbf{1c}^{\mathsf{T}}\right)\left(\mathbf{X}^{\mathsf{T}} - \mathbf{c1}^{\mathsf{T}}\right)\mathbf{m} = \mathbf{0} \tag{32}$$

and, therefore, Equation 30 becomes

$$\mathbf{\Xi S \Xi}^{\mathsf{T}} = \mathbf{S}\,, \tag{33}$$

which lead to

$$-\tfrac{1}{2}\mathbf{\Xi D \Xi}^{\mathsf{T}} = \mathbf{S}\,, \tag{34}$$

which completes the proof.

References

[1] Borg I., & Groenen, P. (1997). *Modern multidimensional scaling.* New York: Springer Verlag.

[2] O'Toole, A.J., Jiang, F., Abdi, H., & Haxby, J.V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience,* **17**, 580–590.

[3] Shepard, R.N. (1966). Metric structure in ordinal data. *Journal of Mathetical Pyschology,* **3**, 287–315.

[4] Togerson, W.S. (1958). *Theory and methods of scaling.* New York: Wiley.