# Inverse Multidimensional scaling

Jan de Leeuw* P.J.F. Groenen†

June 20, 1995

**Abstract**

For metric multidimensional scaling much attention is given to algorithms for computing the configuration for fixed dissimilarities. Here we study the inverse problem: what is the set of dissimilarity matrices that yield a given configuration as a stationary point? Characterisations of this set are given for stationary points, local minima, and for full dimensional scaling. A method for computing the inverse map for stationary points is presented along with several examples.

Keywords: metric multidimensional scaling, inverse map.

## 1 Introduction

The data in a typical multidimensional scaling situation is an $n \times n$ matrix $\Delta = \{\delta_{ij}\}$ of *dissimilarities* between $n$ objects. The dissimilarities are supposed to give imprecise and/or incomplete information about the *distances* of the $n$ objects in some metric space $< X, d >$ . In general terms, the problem is to embed the objects as points in the space in such a way that the distances between the points approximate the dissimilarities between the objects. There are still many variations possible on this theme (cf. [6]). In this paper we restrict our attention to Euclidean scaling, in which $< X, d >$ is a finite-dimensional Euclidean space.

We develop some notation for the Euclidean case. Suppose $X$ are the coordinates of $n$ points in $d$ dimensions. The $n \times d$ matrix $X$ is called a *configuration.* We write $\mathcal{R}^{n\times d}$ for the space of centered configurations (in which the columns of $X$ sum to zero), and we write $d_{ij}(X)$ for the Euclidean distance between points $i$ and $j$.

The basic problem we discuss in this paper is the *Metric Multidimensional Scaling* or MMS problem. In MMS we want to find $X \in \mathcal{R}^{n\times d}$ in such a way

*Interdivisional Program in Statistics, UCLA. Written in part while visiting the National Institute for Statistical Sciences.

†Department of Data Theory, Leiden University.

that the loss function

$$\sigma(X, W, \Delta) \triangleq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(\delta_{ij} - d_{ij}(X))^2 \tag{1.1}$$

is minimized over $X$. Following Kruskal [12],[13] we call $\sigma(X, W, \Delta)$ the `STRESS` of a configuration (for given $W$ and $\Delta$).

We can suppose, without loss of generality, that dissimilarities and weights are symmetric and hollow (have zero diagonal). De Leeuw [2] shows how to partition `STRESS` in such a way that the asymmetric and diagonal parts end up in additive components that do not depend on the configuration. We can also suppose without loss of generality that half the weighted sum of squares of the dissimilarities is equal to one. Moreover, we suppose the weights and dissimilarities are nonnegative. Write $\mathcal{H}^{n\times n}$ for the space of symmetric, nonnegative, and hollow matrices.

The MMS Problem can be made more specific. In order to do this, we have to distinguish between *global minima* and *local minima*.

- A configuration $\hat{X}$ corresponds with a global minimum of `STRESS` if $\sigma(\hat{X}, W, \Delta) \leq \sigma(X, W, \Delta)$ for all $X \in \mathcal{R}^{n\times d}$.
- A configuration $\hat{X}$ corresponds with a local minimum of `STRESS` if there is a neighborhood $\mathcal{N} \subseteq \mathcal{R}^{n\times d}$ of $\hat{X}$ such that $\sigma(\hat{X}, W, \Delta) \leq \sigma(X, W, \Delta)$ for all $X \in \mathcal{N}$.

A problem in MMS is that there are multiple local minima. If local minima were unique, there would be no reason to dinstinguish local minima from global minima in the first place, but all indications [5], [8] are that most MMS problems have a host of different local minima. In order to describe this situation mathematically, we define the (set-valued) maps, on $\mathcal{H}^{n\times n} \times \mathcal{H}^{n\times n}$,

$$\mathcal{X}_{local}(W, \Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid \sigma(X, W, \Delta) \text{ has a local minimum at } X\}, \tag{1.2}$$

$$\mathcal{X}_{global}(W, \Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid \sigma(X, W, \Delta) \text{ has a global minimum at } X\}. \tag{1.3}$$

The first map, the local minima map, associates with each pair $(\Delta, W)$ the configurations that are local minima, the second map does the same with the global minima. MMS can be defined as the technique that studies these local and global minima maps. Any scaling technique is a configuration-valued function that maps data $(W, \Delta)$ into $\mathcal{R}^{n\times d}$, which means that it implements a particular *selection* from the minima-maps. It can be argued that we are really only interested in global minima. Some global minimization techniques for MMS are discussed by Groenen [8], notably the tunneling method (see also Groenen and Heiser [10]). The problems connected with the global minima map have received little attention so far, except in the special case of one-dimensional scaling [11]. Thus we concentrate here on the local minima map, which has been studied in much greater detail, and is a much simpler object. But it helps to think of the local minima map as an approximation of the global minima map. In fact, global minimum algorithms that use multiple random starts use the representation

$$\mathcal{X}_{global}(W,\Delta) = \\ \{\hat{X} \in \mathcal{R}^{n\times d} \mid \sigma(\hat{X},W,D) \leq \sigma(X,W,\Delta) \text{ for all } X \in \mathcal{X}_{local}(W,\Delta)\} \quad (1.4)$$

In this paper we focus on the local and global minimum map for MMS. In particular, we specify what dissimilarity matrices have the fixed configuration $X$ as stationary point and the smaller sets of local minima and global minima. Moreover, we discuss how some of these sets can be computed and give their formal properties. The size of these sets indicates the uniqueness of $\Delta$ for a given $X$. If the set is small then the configuration describes $\Delta$ reliably. Conversely, if the set is large, then $X$ is an unreliable presentation of $\Delta$, since many other $\Delta$ have $X$ as stationary point. We start by describing the maps in more detail.

## 2 Using Differentiability

To study the local minima map we translate some standard results into our notation. Let

$$\mathcal{X}_{diff}(W,\Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid \sigma(X,W,\Delta) \text{ is differentiable at } X\}. \quad (2.1)$$

De Leeuw [3] has shown that if the weights and dissimilarities are non-negative, then

$$\mathcal{X}_{local}(W,\Delta) \subseteq \mathcal{X}_{diff}(W,\Delta). \quad (2.2)$$

But this means that if

$$\mathcal{X}_{partial}(W,\Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid \frac{\partial\sigma(X,W,\Delta)}{\partial X} = 0\}, \quad (2.3)$$

then

$$\mathcal{X}_{local}(W,\Delta) \subseteq \mathcal{X}_{partial}(W,\Delta). \quad (2.4)$$

Most MMS algorithms use gradient or subgradient type methods to find a configuration in $\mathcal{X}_{partial}(W,\Delta)$, and then hope it will also be in $\mathcal{X}_{local}(W,\Delta)$. This is not necessarily true, of course. We can have vanishing partials in saddle points as well (De Leeuw [5] shows that STRESS has no local maxima). Actually we have to be a bit more precise here. The MMS algorithms look for configurations with

$$\left\| \frac{\partial\sigma(X,W,\Delta)}{\partial X} \right\| < \epsilon \quad (2.5)$$

for some small $\epsilon > 0$. If we are in a region where the STRESS is very flat, we still could be a long way from the nearest local minimum (or saddle-point). This makes it necessary, in practice, to look at the second derivatives of STRESS as well.

The second partials make it possible to make (2.4) more precise. We define the regions where the Hessian is non-negative definite, and where it is positive definite. We write them as

$$\mathcal{X}_{nne-hes}(W,\Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid \frac{\partial^2\sigma(X,W,\Delta)}{\partial X^2} \gtrsim 0\}, \tag{2.6}$$

$$\mathcal{X}_{pos-hes}(W,\Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid \frac{\partial^2\sigma(X,W,\Delta)}{\partial X^2} \gtrdot 0\}. \tag{2.7}$$

It follows that

$$\begin{aligned}\mathcal{X}_{pos-hes}(W,\Delta) \cap \mathcal{X}_{partial}(W,\Delta) &\subseteq \mathcal{X}_{local}(W,\Delta)\\ &\subseteq \mathcal{X}_{nne-hes}(W,\Delta) \cap \mathcal{X}_{partial}(W,\Delta).\end{aligned} \tag{2.8}$$

This is just saying that a necessary condition for a configuration to be a local minimum is that the partials vanish and the Hessian is non-negative definite, a sufficient condition is that the partials vanish and the Hessian is positive definite. Let

$$\mathcal{X}_{l-local}(W,\Delta) \triangleq \mathcal{X}_{pos-hes}(W,\Delta) \cap \mathcal{X}_{partial}(W,\Delta), \tag{2.9}$$

$$\mathcal{X}_{u-local}(W,\Delta) \triangleq \mathcal{X}_{nne-hes}(W,\Delta) \cap \mathcal{X}_{partial}(W,\Delta). \tag{2.10}$$

Then, instead of studying $\mathcal{X}_{local}$ directly, we can study $\mathcal{X}_{partial}$ or $\mathcal{X}_{l-local}$ and $\mathcal{X}_{u-local}$. These maps are far from simple. De Leeuw [5] has shown that STRESS has local minima, sharp ridges, and other irregularities. There seems to be no obvious relationship between the different local minima, and there are no systematic results on the number of local minima. In order to compute the map, or a selection from the map, we need complicated iterative algorithms, perhaps with multiple random starts. Some results are available for very special cases, such as unidimensional scaling and full-dimensional scaling (cf. below), but for $1 < p < n-1$ almost nothing is known.

# 3 Inverse Metric Multidimensional Scaling

In order to understand the mappings $\mathcal{X}_{partial}$, $\mathcal{X}_{l-local}$, and $\mathcal{X}_{u-local}$ a bit better, we look at their inverses. Thus in stead of finding the configurations which are optimal for a given set of weights and dissimilarities, we now look at the weights and dissimilarities for which a given configuration is optimal. There is one obvious reason to do this: it turns out that the inverse maps are comparatively simple. And by studying the inverses in detail, we learn a great deal about the maps themselves. There is a useful analogous situation. In an eigenvalue problem we compute the eigenvectors of a given matrix, in an inverse eigenvalue problem we compute matrices of which a given orthogonal system is a matrix of eigenvectors. MMS is quite close to an eigenvalue problem in various aspects [2], although versions of MMS that use SSTRESS or STRAIN are much more like eigenvalue problems. The inverse MMS problem for SSTRESS and STRAIN is discussed in Groenen, De Leeuw, and Mathar [9].

For the time being, we restrict ourselves to configurations $X$ which have $d_{ij} > 0$ for all $i \neq j$. Since we are interested in local minima, this causes no real loss of generality [3]. The inverse of $\mathcal{X}_{partial}$, for instance, is defined as

$$\mathcal{X}^{+}_{partial}(X) \triangleq \{W \in \mathcal{H}^{n\times n}, \Delta \in \mathcal{H}^{n\times n} \mid \frac{\partial\sigma(X,W,\Delta)}{\partial X} = 0\}. \tag{3.1}$$

Inverses for the other maps are defined in the same way, but we will analyze the partial-map in this section. In order to do that efficiently we also define

$$\mathcal{X}^{+}_{partial}(X,W) \triangleq \{\Delta \in \mathcal{H}^{n\times n} \mid \frac{\partial\sigma(X,W,\Delta)}{\partial X} = 0\}. \tag{3.2}$$

This is just the set of dissimilarity matrices for which $X$ is stationary for given $W$. For our computations, we also need an orthonormal column-centered matrix $K$, of dimensions $n \times (n-r-1)$, such that $K'X = 0$. Here $r \triangleq \text{rank}(X)$.

**Theorem 3.1 (Inverse).**

$$\mathcal{X}^{+}_{partial}(X,W) = \{\Delta \in \mathcal{H}^{n\times n} \mid \delta_{ij} = d_{ij}(1 - \frac{t_{ij}}{w_{ij}})\}, \tag{3.3}$$

*where $T$ is of the form $T = KMK'$, with $M$ an arbitrary real symmetric matrix (of order $n-r-1$), and satisfies $t_{ij} \leq w_{ij}$ for all $i \neq j$.*

*Proof.* The stationary equations have to be brought into a convenient form. We use the notation familiar from earlier papers, such as [6], which gives

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij}X = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{\delta_{ij}}{d_{ij}}A_{ij}X, \tag{3.4}$$

with

$$A_{ij} \triangleq (e_i - e_j)(e_i - e_j)', \tag{3.5}$$

and with the $e_i$ the unit-vectors of $\mathcal{R}^n$. We have to solve (3.4) for $\Delta$ for given $X$ and $W$. We make the transformation indicated in the Theorem, i.e. we define

$$t_{ij} \triangleq w_{ij} - w_{ij}\frac{\delta_{ij}}{d_{ij}}, \tag{3.6}$$

and we solve for $t_{ij}$. Equation (3.4) transforms to

$$\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij}A_{ij}X = 0. \tag{3.7}$$

To solve (3.7) we have to realize that the $A_{ij}$ are a basis for the symmetric, doubly centered (SDC) matrices of order $n$. Thus (3.7) is solved if we find all SDC matrices $T$ such that $TX = 0$. But that means $T = KMK'$, with $M$ an arbitrary symmetric matrix. Thus there are $\frac{1}{2}(n-r)(n-r-1)$ independent solutions in all.

Of course we must also have $\delta_{ij} \geq 0$, which translates to $t_{ij} \leq w_{ij}$. $\square$

A brief comment is in order here. The $t_{ij}$ are defined by (3.6) only for $i \neq j$. We can define the $t_{ii}$ in a completely arbitrary way, because no matter how we define them (3.7) will still be true. Thus (3.7) does not define the $t_{ii}$ and we simply choose them in such a way that $T$ is SDC.

In order to facilitate comparison with other basic MMS papers, such as [4], [6], [7], we define

$$V \triangleq \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij}, \tag{3.8}$$

$$B \triangleq \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{\delta_{ij}}{d_{ij}}A_{ij}. \tag{3.9}$$

In this notation $T = V - B$, and Theorem Inverse simply says that $B = V - KMK'$.

**Corollary 3.2 (Bounded).** *$\mathcal{X}^{+}_{partial}(X, W)$ is a closed, bounded, convex polyhedron, containing $D(X)$.*

*Proof.* Closedness, polyhedrality, and convexity follow directly from the representation in the theorem. Obviously $D(X) \in \mathcal{X}^{+}_{partial}(X, W)$. Only boundedness is nontrivial. We have to show [15] that the set cannot contain a ray. From (3.7) we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij}d^2_{ij}(X) = 0. \tag{3.10}$$

This means that not all $t_{ij}$ can have the same sign, at least one $t_{ij}$ has to be negative. For this $t_{ij}$, and for some $\lambda > 0$ we have that $\lambda t_{ij} < -w_{ij}$, and thus the set of matrices $T$ cannot be unbounded. $\square$

**Corollary 3.3 (Dominate).** *If $\Delta_1 \in \mathcal{X}^{+}_{partial}(X, W)$ and $\Delta_2 \in \mathcal{X}^{+}_{partial}(X, W)$ and $\delta_{ij1} \leq \delta_{ij2}$ for all $i < j$, then actually $\Delta_1 = \Delta_2$.*

*Proof.* We have $\delta_{ij1} \leq \delta_{ij2}$ if and only if $t_{ij1} \leq t_{ij2}$. But, from (3.10),

$$\sum_{i=1}^{n}\sum_{j=1}^{n} (t_{ij1} - t_{ij2})d^2_{ij}(X) = 0, \tag{3.11}$$

which is impossible unless $T_1 = T_2$. $\square$

**Corollary 3.4 (Only).**

$$\mathcal{X}^+_{partial}(X) = \{W \in \mathcal{H}^{n\times n}, \Delta \in \mathcal{H}^{n\times n} \mid \Delta \in \mathcal{X}^+_{partial}(X,W)\}. \tag{3.12}$$

*Proof.* Directly from the representation in the Theorem. $\square$

From the last corrollary we can choose $W$ arbitrarily in $\mathcal{H}^{n\times n}$, and for each $W$ there is a corresponding set of dissimilarities. Thus weights are not very essential to the formulation of the problem, and we shall largely ignore them from now on.

If $\Delta_1$ and $\Delta_2$ are two different elements of $\mathcal{X}^+_{partial}$, then we can measure their distance by using

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(\delta_{1ij} - \delta_{2ij})^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d^2_{ij}(t_{ij1} - t_{ij2})^2. \tag{3.13}$$

In particular, the distance between $\Delta$ and $D(X)$, which of course is simply `STRESS`, is equal to

$$\sigma(X,W,\Delta) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d^2_{ij}t^2_{ij}. \tag{3.14}$$

Now suppose we have $m$ configurations $X_1, \cdots, X_m$. We can ask for the set of dissimilarity matrices for which all $X_j$ are stationary points. It is not necessary that all $X_j$ have the same rank. Each of the configurations defines an affine space of dimension $\frac{1}{2}(n-r_j)(n-r_j-1)$, and if these spaces are "in general position", they have an intersection if $\frac{1}{2}\sum_{j=1}^{m}(n-r_j)(n-r_j-1) \geq \frac{1}{2}(m-1)n(n-1)$. If all $r_j$ are equal to $p$ this works out to

$$m \leq \frac{n(n-1)}{n(n-1)-(n-p)(n-p-1)}.. \tag{3.15}$$

It is tempting to use (3.15) as an upper bound on the number of stationary points of *STRESS,* but the reasoning here is difficult to make rigourous.

# 4 Computing the Inverse Map

We now go into more detail in describing the convex polyhedron defined in Theorem Inverse. From the computational point of view, it is convenient to use a basis $P_r$ for the symmetric matrices of order $n-r-1$. Define

$$Q_r \triangleq KP_rK', \tag{4.1}$$

and then write

$$T = \sum_{r=1}^{R} \theta_r Q_r. \tag{4.2}$$

If we limit ourselves to the case $w_{ij} = 1$ for all $i \neq j$ then we must have

$$\sum_{r=1}^{R} \theta_r q_{rij} \leq 1 \tag{4.3}$$

for all $i < j$, which are $N \triangleq \frac{1}{2}n(n-1)$ linear inequalities in $R$ unknowns. Obviously, these linear inequalities describe the bounded convex polyhedron of Corollary Bounded.

Bounded convex polyhedra can be described in term of their edges. Compare [14], [16]. We find the edges of the polyhedron by an enumerative procedure which looks at all subsystems of $R$ rows of (4.3). If the complete system is written as $Q\theta \geq -u$, then we can write a subsystem as

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \theta \geq - \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{4.4}$$

with $Q_1$ of order $R$. We then check if $Q_1$ is nonsingular. If it is not, we go to the next subsystem. If it is, we compute $\tilde{\theta} = -Q_1^{-1}u_1$. If $Q_2\tilde{\theta} \geq -u_2$ we add $\tilde{\theta}$ to our list of edges. If not, we go to the next subsystem. This can be done quite efficiently by using pivoting techniques, moving one row into the basis and another one out of the basis in one pivot, and cycling through the candidate subsets lexicographically [1]. In the examples below we simply use brute force, and investigate all subsets.

We start with a really simple example. Let's call it Example Square. Consider the configuration

$$X = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ +\frac{1}{2} & -\frac{1}{2} \\ +\frac{1}{2} & +\frac{1}{2} \\ -\frac{1}{2} & +\frac{1}{2} \end{pmatrix}, \text{ with distances } D = \begin{pmatrix} 0 & 1 & \sqrt{2} & 1 \\ 1 & 0 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 0 & 1 \\ 1 & \sqrt{2} & 1 & 0 \end{pmatrix}.$$

We now want to find all dissimilarity matrices $\Delta$ for which $X$ gives a stationary value of STRESS. Throughout, we fix $W$ at $w_{ij} = 1$. For $K$ we find

$$K = \begin{pmatrix} -\frac{1}{2} \\ +\frac{1}{2} \\ -\frac{1}{2} \\ +\frac{1}{2} \end{pmatrix}, \text{ and thus } T = \theta \begin{pmatrix} +1 & -1 & +1 & -1 \\ -1 & +1 & -1 & +1 \\ +1 & -1 & +1 & -1 \\ -1 & +1 & -1 & +1 \end{pmatrix},$$

with $-1 \leq \theta \leq +1$. The two extreme points are

$$\Delta_1 = \begin{pmatrix} 0 & 0 & 2\sqrt{2} & 0 \\ 0 & 0 & 0 & 2\sqrt{2} \\ 2\sqrt{2} & 0 & 0 & 0 \\ 0 & 2\sqrt{2} & 0 & 0 \end{pmatrix}, \text{ and } \Delta_2 = \begin{pmatrix} 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \end{pmatrix}.$$

By defining $\lambda = (1-\theta)/2$, we can say that any convex combination $\Delta(\lambda) = \lambda\Delta_1 + (1-\lambda)\Delta_2$ of these edges has the four points arranged on a "square", with the length of the side equal to $2(1-\lambda)$ and the length of the diagonals equal to $2\lambda\sqrt{2}$. It follows directly from this interpretation that for $\Delta(\lambda)$ to be Euclidean we need to have $\lambda \leq \frac{1}{2}$, while $\Delta(\lambda)$ statisfies all triangle inequalities for $\lambda \leq 2/(2+\sqrt{2}) \approx .586$. The distance matrix $D = \Delta(\frac{1}{2})$ is exactly in the middle of the edges. In the interval we also have the matrix with all six dissimilarities equal, which is the distance matrix of a regular simplex in three dimensions. For this matrix $\lambda = 1/(1+\sqrt{2}) \approx .414$. For the distances between the edges and their centroid $D$ we find

$$\begin{array}{c c} & \begin{array}{ccc} D & \Delta_1 & \Delta_2 \end{array} \\ \begin{array}{c} D \\ \Delta_1 \\ \Delta_2 \end{array} & \begin{pmatrix} 0 & 8 & 8 \\ 8 & 0 & 32 \\ 8 & 32 & 0 \end{pmatrix} \end{array}.$$

Thus the STRESS of both edges is 8.

# 5 Improved Approximation

We know that $\mathcal{X}^{+}_{partial}(X,W)$ is a compact convex set. It is clear from Equation 2.8 that $\mathcal{X}_{u-local}(X,W)$ is a more precise approximation of $\mathcal{X}_{local}(X,W)$, and we shall see that its inverse also is convex and compact (although not necessarily polyhedral).

First, we need a convenient expression for the Hessian of STRESS. This has been discussed earlier in [4]. We start with

$$\frac{\partial\sigma}{\partial x_s} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij}x_s - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{\delta_{ij}}{d_{ij}}A_{ij}x_s. \tag{5.1}$$

Thus

$$\begin{aligned} \frac{\partial^2\sigma}{\partial x_s\partial x_t} &= \delta^{st}\{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{\delta_{ij}}{d_{ij}}A_{ij}\} \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{\delta_{ij}}{d_{ij}^3}A_{ij}x_s x_t' A_{ij}. \end{aligned} \tag{5.2}$$

Now substitute

$$w_{ij}\frac{\delta_{ij}}{d_{ij}} = w_{ij} - t_{ij}. \tag{5.3}$$

This gives

$$\begin{aligned} \frac{\partial^2\sigma}{\partial x_s\partial x_t} &= \delta^{st}T - \sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij}\frac{(x_{is}-x_{js})(x_{it}-x_{jt})}{d_{ij}^2}A_{ij} + \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{(x_{is}-x_{js})(x_{it}-x_{jt})}{d_{ij}^2}A_{ij}. \end{aligned} \tag{5.4}$$

Here superscripted $\delta$ is the Kronecker symbol, it does not have anything to do with the subscripted $\delta$'s. It is convenient at this point to define the $np \times np$ supermatrices $H_0, H_1, \cdots, H_R$ with submatrices

$$H_{rst} \triangleq \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ijr}\frac{(x_{is}-x_{js})(x_{it}-x_{jt})}{d_{ij}^2}A_{ij}, \tag{5.5}$$

$$H_{0st} \triangleq \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{(x_{is}-x_{js})(x_{it}-x_{jt})}{d_{ij}^2}A_{ij}, \tag{5.6}$$

where $Q_r$ is given by Equation 4.1. Also

$$\overline{Q}_r \triangleq \underbrace{Q_r \oplus \cdots \oplus Q_r}_{\text{p times}}, \tag{5.7}$$

i.e. $\overline{Q}_r$ is the diagonal supermatrix with the $Q_r$ repeated along the diagonal. Using these definitions, we find

$$\frac{\partial^2\sigma}{\partial x \partial x} = H_0 + \sum_{r=1}^{R}\theta_r(\overline{Q}_r - H_r). \tag{5.8}$$

**Theorem 5.1 (Improved).** *$\mathcal{X}^+_{u-local}(X,W)$ is a compact convex set.*

*Proof.* We have $\Delta \in \mathcal{X}^+_{u-local}(X,W)$ if and only if $\Delta$ is of the form in Theorem Inverse, with in addition

$$\sum_{r=1}^{R}\theta_r(\overline{Q}_r - H_r) \gtrsim -H_0. \tag{5.9}$$

But this means that $\mathcal{X}^+_{u-local}(X,W)$ is the intersection of the convex set defined by (5.9) and the compact convex set from Theorem Inverse, i.e. it is a compact convex set. $\square$

Unfortunately, $\mathcal{X}^+_{u-local}(X,W)$ is more difficult to describe than $\mathcal{X}^+_{partial}(X,W)$, because it is not polyhedral. We can approximate it by polyhedral sets, by cutting off the edges that are not in the cone, using the eigenvectors corresponding to the positive eigenvalues. This makes arbitrarily precise approximation possible, but the number of edges will increase very rapidly.

In our Example Square, we can still carry out the necessary computations quite easily. We know from the results of De Leeuw [4] that the Hessian has at least $\frac{1}{2}p(p+1)$ eigenvalues equal to zero, corresponding with the rotational and translational invariance of the distances, and it has at least one eigenvalue equal to $n$. In the example, the smallest eigenvalue is equal to zero for $-\frac{1}{2} \leq \theta \leq 1$, and it is negative for $-1 \leq \theta \leq -\frac{1}{2}$, i.e. $3/4 \leq \lambda \leq 1$. Thus the more precise approximation tells us that for a local minimum we must have $0 \leq \lambda \leq 3/4$. At $\lambda = 3/4$, the STRESS is 2, and

$$\Delta = \begin{pmatrix} 0 & \frac{1}{2} & \frac{3}{2}\sqrt{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{3}{2}\sqrt{2} \\ \frac{3}{2}\sqrt{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2}\sqrt{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

More examples of inverse scaling are given in Appendix A.

# 6 Full Dimensional Scaling

In MDS we minimize STRESS over $D$, on the conditon that $D = D(X)$, i.e. $D$ are the Euclidean distances between the $n$ points of a configuration in $p$ dimensions. Now suppose we drop the constraint of $p$ dimensions, and merely require that the $D$ are Euclidean distances between points of any configuration. This defines *full dimensional scaling,* or FDS. There is no need to emphasize the fact that FDS is metric, because non-metric full-dimensional scaling does not make sense (any dissimilarity matrix can be fitted perfectly in $n-2$ dimensions nonmetrically). The most interesting result on FDS is that all local minima are global. This result is due to De Leeuw [5], but because the proof is difficult to find and simple to reproduce, we give it here for completeness.

**Theorem 6.1 (Full).** *In FDS all local minima are global.*

*Proof.* The FDS problem can be formulated as minimization of

$$\sigma(C) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(\delta_{ij} - d_{ij}(C))^2 \tag{6.1}$$

over the convex cone of all positive semi-definite matrices $C$. Here

$$d_{ij}(C) = \sqrt{c_{ii} + c_{jj} - 2c_{ij}}. \tag{6.2}$$

Thus $d_{ij}(C)$ is the square root of a linear function of $C$, which means it is concave in $C$. Obviously $d_{ij}^2(C)$ is linear in $C$. It follows that $\sigma(C)$ is convex in $C$, and thus the FDS problem minimizes a convex function over a convex set. All local mimima are global. $\square$

It now makes sense to define *inverse* FDS. Given a configuration $X$, find the weights and/or dissimilarities for which $X$ is the unique solution to the FDS problem. Thus we define

$$\mathcal{X}_{full}(W,\Delta) \triangleq \{X \in \mathcal{R}^{n\times d} \mid X \text{solves the FDS problem}\}, \tag{6.3}$$

$$\mathcal{X}_{full}^{+}(W,\Delta) \triangleq \{\Delta \in \mathcal{H}^{n\times n} \mid X \text{solves the FDS problem}\}. \tag{6.4}$$

**Theorem 6.2 (Inverse Full).** *$\mathcal{X}^+_{full}(W, \Delta)$ is a compact convex set.*

*Proof.* If we minimize a differentiable convex function $f(\bullet)$ over a convex cone $\mathcal{K}$, then the necessary and sufficient conditions for a minimum [15] are

- $\hat{x} \in \mathcal{K}$,
- $-\nabla f(\hat{x}) \in \mathcal{K}^\circ$,
- $< \hat{x}, \nabla f(\hat{x}) >= 0$.

In our case this means that the necessary and sufficient conditions for the FDS problem are

$$C \gtrsim 0, \tag{6.5}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{w_{ij}\delta_{ij}}{d_{ij}(C)}A_{ij} \gtrsim 0, \tag{6.6}$$

$$\text{tr } C\Big\{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}A_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{w_{ij}\delta_{ij}}{d_{ij}(C)}A_{ij}\Big\} = 0, \tag{6.7}$$

which translates to $T \gtrsim 0$ and $TX = 0$. But this is the same as $T = KMK'$, with $M$ a positive semidefinite matrix. So again we have the intersection of a convex cone and the compact convex set of Theorem Inverse. □

For our simple example we have $T \gtrsim 0$ if and only if $\theta \geq 0$. Thus the dissimilarity matrices for which $X$ solves the FDS problem are the ones on the line segment between $D$ and $\Delta_2$.

## References


[1] DANTZIG, G.B, (1963), "Linear Programming and Extensions", Princeton, New Jersey: Princeton University Press.

[2] DE LEEUW, J. (1977), "Applications of convex analysis to multidimensional scaling," in *Recent developments in statistics*, Eds., J.R. Barra, F. Brodeau, G. Romier, and B. van Cutsem, Amsterdam: North-Holland, 133-145.

[3] DE LEEUW, J. (1984), "Differentiability of Kruskal's Stress at a local minimum," *Psychometrika, 49*, 111-113.

[4] DE LEEUW, J. (1988), "Convergence of the majorization method for multidimensional scaling," *Journal of Classification*, *5*, 163-180.

[5] DE LEEUW, J. (1993), *Fitting distances by least squares*, Technical Report, UCLA Statistics Series #130, Los Angeles: UCLA.

[6] DE LEEUW, J., and HEISER, W.J. (1980), "Multidimensional scaling with restrictions on the configuration," in *Multivariate analysis V*, Ed., P.R. Krishnaiah, Amsterdam: North-Holland, 501-522.

[7] DE LEEUW, J., and HEISER, W.J. (1982), "Theory of multidimensional scaling", in *Handbook of statistics, vol 2*, Ed., P.R. Krishnaiah, Amsterdam: North-Holland, 285-316.

[8] GROENEN, P.J.F. (1993) *The majorization approach to multidimensional scaling: some problems and extensions*, Leiden: DSWO Press.

[9] GROENEN, P.J.F., DE LEEUW, J., and MATHAR, R. (in press), "Least squares multidimensional scaling with transformed distances", in *Studies in classification, data analysis, and knowledge organization*, Eds., W. Gaul and D. Pfeifer, Berlin: Springer, 000-000.

[10] GROENEN, P.J.F., and HEISER, W.J. (in press). "The tunneling method for global optimization in multidimensional scaling", *Psychometrika, 00*, 00-00.

[11] HUBERT, L.J., and ARABIE, P. (1986), "Unidimensional scaling and combinatorial optimization," in *Multidimensional Data Analysis*, Eds., J. De Leeuw, W.J. Heiser, J. Meulman, and F. Critchley, Leiden: DSWO Press, 181-196.

[12] KRUSKAL, J.B. (1964a), "Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis," *Psychometrika, 29*, 1-27.

[13] KRUSKAL, J.B. (1964b), "Nonmetric multidimensional scaling: a numerical method," *Psychometrika, 29*, 115-129.

[14] KUHN, H.W., and TUCKER, A.W. (1956), “Linear inequalities and related systems”, Volume 38 of *Annals of Mathematics Studies*, Princeton, New Jersey: Princeton University Press.

[15] ROCKAFELLAR, R.T. (1970), *Convex Analysis*, Princeton, New Jersey: Princeton University Press.

[16] TSCHERNIKOW, S.N. (1971), *Lineare Ungleichungen*, Berlin, GDR: VEB Deutscher Verlag der Wissenschaften.

# A Appendix

In this appendix two examples of inverse scaling are reported in detail and the results of four other examples are summarized in a table.

The first example concerns a configuration of four points equally spaced on a line. The coordinates are

$$X_l = \begin{pmatrix} -3 \\ -1 \\ +1 \\ +3 \end{pmatrix}, \text{ which has null space } K_l = \begin{pmatrix} +2 & +1 \\ -2 & -3 \\ -2 & +3 \\ +2 & -1 \end{pmatrix}.$$

We have found seven vertices of that produces dissimilarities with $X_l$ as local minima. They are local minima, because the STRESS is a piecewise linear quadratic function, where the pieces depends only on the order of the coordinates of $X$. The vertices are summarized in Table 1.

Insert Table 1 about here.

The second example consists of the square configuration discussed in Section 'Computing the inverse map' extended by a point in the centroid. The configuration becomes

$$X_s = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ +\frac{1}{2} & -\frac{1}{2} \\ +\frac{1}{2} & +\frac{1}{2} \\ -\frac{1}{2} & +\frac{1}{2} \\ 0 & 0 \end{pmatrix}, \text{ which has null space } K_s = \begin{pmatrix} +\frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ -\frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ +\frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ -\frac{1}{2} & -\frac{1}{2\sqrt{5}} \\ 0 & +\frac{2}{\sqrt{5}} \end{pmatrix}.$$

Five vertices were obtained with inverse scaling. The vertices and some of their properties are described in Table 2.

Insert Table 2 about here.

The results of inverse scaling of four additional examples are given in Table 3.

Insert Table 3 about here .

Table 1: The vertices of the polyhedral set that defines dissimilarities $\Delta$ for which $X_l$ is a local minimum.

| vertex | $M$ | STRESS | $\Delta$ |
|---|---|---|---|
| 1 | $\begin{pmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{6} \end{pmatrix}$ | 16 | $\begin{pmatrix} 0 & 0 & 4 & 8 \\ 0 & 0 & 0 & 4 \\ 4 & 0 & 0 & 0 \\ 8 & 4 & 0 & 0 \end{pmatrix}$ |
| 2 | $\begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{pmatrix}$ | 80 | $\begin{pmatrix} 0 & 0 & 0 & 12 \\ 0 & 0 & 4 & 0 \\ 0 & 4 & 0 & 0 \\ 12 & 0 & 0 & 0 \end{pmatrix}$ |
| 3 | $\begin{pmatrix} -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 \end{pmatrix}$ | 144 | $\begin{pmatrix} 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 4 \\ 12 & 0 & 0 & 8 \\ 0 & 4 & 8 & 0 \end{pmatrix}$ |
| 4 | $\begin{pmatrix} -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & 0 \end{pmatrix}$ | 144 | $\begin{pmatrix} 0 & 8 & 4 & 0 \\ 8 & 0 & 0 & 12 \\ 4 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 \end{pmatrix}$ |
| 5 | $\begin{pmatrix} -\frac{5}{16} & \frac{3}{8} \\ \frac{3}{8} & -\frac{1}{4} \end{pmatrix}$ | 224 | $\begin{pmatrix} 0 & 0 & 12 & 0 \\ 0 & 0 & 4 & 0 \\ 12 & 4 & 0 & 12 \\ 0 & 0 & 12 & 0 \end{pmatrix}$ |
| 6 | $\begin{pmatrix} -\frac{5}{16} & -\frac{3}{8} \\ -\frac{3}{8} & -\frac{1}{4} \end{pmatrix}$ | 224 | $\begin{pmatrix} 0 & 12 & 0 & 0 \\ 12 & 0 & 4 & 12 \\ 0 & 4 & 0 & 0 \\ 0 & 12 & 0 & 0 \end{pmatrix}$ |
| 7 | $\begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & -1 \end{pmatrix}$ | 464 | $\begin{pmatrix} 0 & 12 & 0 & 0 \\ 12 & 0 & 16 & 0 \\ 0 & 16 & 0 & 12 \\ 0 & 0 & 12 & 0 \end{pmatrix}$ |

Table 2: The vertices of the polyhedral set that defines dissimilarities $\Delta$ for which $X_s$ is a stationary point. $\lambda_i$ is the $i$-th eigenvalue of the Hessian $H$.

| vertex | $M$ | STRESS | $\Delta$ | Eigenvalues of $H$ |
|---|---|---|---|---|
| 1 | $\begin{pmatrix} 0 & 0 \\ 0 & 20 \end{pmatrix}$ | 40 | $\begin{pmatrix} 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ \frac{5}{\sqrt{2}} & \frac{5}{\sqrt{2}} & \frac{5}{\sqrt{2}} & \frac{5}{\sqrt{2}} & 0 \end{pmatrix}$ | $\lambda_1 = 5$, $\lambda_6 = 0$, $\lambda_2 = 5$, $\lambda_7 = 0$, $\lambda_3 = 5$, $\lambda_8 = 0$, $\lambda_4 = 5$, $\lambda_9 = -10$, $\lambda_5 = 0$, $\lambda_{10} = -10$ |
| 2 | $\begin{pmatrix} -\frac{5}{2} & -\frac{5}{2}\sqrt{5} \\ -\frac{5}{2}\sqrt{5} & \frac{15}{2} \end{pmatrix}$ | 27.5 | $\begin{pmatrix} 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ 0 & 0 & 0 & \frac{5}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ 0 & \frac{5}{\sqrt{2}} & 0 & 0 & 0 \\ \frac{5}{\sqrt{2}} & 0 & \frac{5}{\sqrt{2}} & 0 & 0 \end{pmatrix}$ | $\lambda_1 = 5$, $\lambda_6 = 0$, $\lambda_2 = 5$, $\lambda_7 = 0$, $\lambda_3 = 5$, $\lambda_8 = 0$, $\lambda_4 = 5$, $\lambda_9 = 0$, $\lambda_5 = 5$, $\lambda_{10} = -10$ |
| 3 | $\begin{pmatrix} -\frac{5}{2} & \frac{5}{2}\sqrt{5} \\ \frac{5}{2}\sqrt{5} & \frac{15}{2} \end{pmatrix}$ | 27.5 | $\begin{pmatrix} 0 & 0 & \frac{5}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ \frac{5}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{\sqrt{2}} \\ \frac{5}{\sqrt{2}} & 0 & \frac{5}{\sqrt{2}} & 0 & 0 \end{pmatrix}$ | $\lambda_1 = 5$, $\lambda_6 = 0$, $\lambda_2 = 5$, $\lambda_7 = 0$, $\lambda_3 = 5$, $\lambda_8 = 0$, $\lambda_4 = 5$, $\lambda_9 = 0$, $\lambda_5 = 5$, $\lambda_{10} = -10$ |
| 4 | $\begin{pmatrix} -5 & 0 \\ 0 & -5 \end{pmatrix}$ | 15 | $\begin{pmatrix} 0 & 0 & \frac{5}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & \frac{5}{\sqrt{2}} & 0 \\ \frac{5}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & \frac{5}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ | $\lambda_1 = 5$, $\lambda_6 = 5$, $\lambda_2 = 5$, $\lambda_7 = 0$, $\lambda_3 = 5$, $\lambda_8 = 0$, $\lambda_4 = 5$, $\lambda_9 = 0$, $\lambda_5 = 5$, $\lambda_{10} = 0$ |
| 5 | $\begin{pmatrix} 5 & 0 \\ 0 & -5 \end{pmatrix}$ | 15 | $\begin{pmatrix} 0 & \frac{5}{2} & 0 & \frac{5}{2} & 0 \\ \frac{5}{2} & 0 & \frac{5}{2} & 0 & 0 \\ 0 & \frac{5}{2} & 0 & \frac{5}{2} & 0 \\ \frac{5}{2} & 0 & \frac{5}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ | $\lambda_1 = 5$, $\lambda_6 = 0$, $\lambda_2 = 5$, $\lambda_7 = 0$, $\lambda_3 = 5$, $\lambda_8 = 0$, $\lambda_4 = 5$, $\lambda_9 = 0$, $\lambda_5 = 0$, $\lambda_{10} = 0$ |

Table 3: Six configurations and the results of their vertices obtained by inverse scaling.

| | Example | # vertices | # vertices for which $X$ is a local minimum | # vertices for which $X$ is a fulldimensional scaling solution |
|---|---|---|---|---|
| a. | four points equally spaced on a line | 7 | 7 | 1 |
| b. | equilaterial triangle with centroid | 2 | 1 | 0 |
| c. | square | 2 | 1 | 1 |
| d. | square with centroid | 5 | 2 | 0 |
| e. | five points equally spaced on a circle | 7 | 6 | 0 |
| f. | six points equally spaced on a circle | 42 | 9 | 0 |