Journal of Classification 14:3-21 (1997)

Inverse Multidimensional Scaling

Jan de Leeuw
University of California at Los Angeles

Patrick J. F. Groenen
Leiden University

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? Characterizations 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 (MDS) situation are 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* among the n objects in some metric space <**X,d**>. In general

---

We would like to thank two anonymous referees for carefully reading the manuscript and for several valuable remarks that have improved this paper.

Authors' addresses: Jan de Leeuw, University of California Los Angeles, UCLA Department of Statistics, 8118 Math Sciences, 405 Hilgard Avenue, Los Angeles, CA 90095-1554, USA (e-mail: deleeuw@stat.ucla.edu). Written in part while visiting the National Institute for Statistical Sciences. Patrick J. F. Groenen, Department of Data Theory, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands (e-mail: groenen@rulfsw.fsw.leidenuniv.nl). Supported by The Netherlands Organization for Scientific Research (NWO) by grant nr. 575-67-053 for the 'PIONEER' project 'Subject Oriented Multivariate Analysis'.

terms, the problem is to embed the objects as points in the space in such a way that the distances between the pairs of points approximate the dissimilarities between distinct pairs of the objects. There are still many variations possible on this theme (cf. De Leeuw and Heiser 1980). In this paper we restrict our attention to Euclidean scaling, in which $<\mathbf{X},\mathbf{d}>$ is a finite-dimensional Euclidean space.

We develop some notation for the Euclidean case. Suppose $\mathbf{X}$ contains the coordinates of n points in p dimensions. The $n \times p$ matrix $\mathbf{X}$ is called a *configuration*. We write $R^{n \times p}$ for the space of centered configurations (in which each column of $\mathbf{X}$ sums to zero), and we write $d_{ij}(\mathbf{X})$ (or d_{ij} for short) for the Euclidean distance between points i and j.

The basic problem we discuss in this paper is the *Metric Multidimensional Scaling* or MMDS problem. In MMDS we want to find $\mathbf{X} \in R^{n \times p}$ so that the loss function

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

is minimized over $\mathbf{X}$. We suppose that the weights w_{ij} and dissimilarities δ_{ij} are nonnegative. Following Kruskal (1964a, 1964b) we call $\sigma(\mathbf{X},\mathbf{W},\Delta)$ the STRESS of a configuration (for given $\mathbf{W}$ and Δ). (Actually, Kruskal uses the square-root of (1.1), to simplify the interpretation of numerical STRESS values.)

We can suppose, without loss of generality, that dissimilarities and weights are symmetric and have zero diagonal, because it is possible 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 (see De Leeuw 1977). Write $H^{n \times n}$ for the space of symmetric, nonnegative matrices with zero diagonal.

The MMDS problem can be made more specific. To do so, we have to distinguish between *global minima* and *local minima.*

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

It was realized quite early (Kruskal 1964b, p. 118-119) that MMDS has multiple local minima. If those local minima were unique, there would be no reason to distinguish local minima from global minima in the first place, but all indications (De Leeuw 1993; Groenen 1993) are that most MMDS problems have a host of different local minima. To describe this situation mathematically, we define the (set-valued) maps, on $H^{n\times n} \times H^{n\times n}$,

$$X_{local}(\mathbf{W},\Delta) \triangleq \{\mathbf{X} \in R^{n\times p} \mid \sigma(\mathbf{X},\mathbf{W},\Delta) \text{ has a local minimum at } \mathbf{X}\} ; \tag{1.2}$$

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

The first map, the local minimum map, associates with each pair $(\Delta,\mathbf{W})$ the configurations that are local minima; the second map does the same with the global minima. MMDS can be defined as the technique that studies these local and global minimum maps. Any MDS technique is a configuration-valued function that maps data $(\mathbf{W},\Delta)$ into $R^{n\times p}$, which means that it implements a particular *selection* from the minimum-maps. It can be argued that we are really only interested in global minima. Some global minimization techniques for MMDS are discussed by Groenen (1993), notably the tunneling method (see also Groenen and Heiser, 1996). However, the problems connected with the global minimum map have received little attention so far, except in the special case of unidimensional scaling (Hubert and Arabie 1986). Thus, we concentrate here on the local minimum map, which has been studies in much greater detail, and is a much simpler object. But it helps to think of the local minimum map as an approximation of the global minimum map. In fact, global minimum algorithms that use multiple random starts use the representation

$$X_{global}(\mathbf{W},\Delta) = \{\hat{\mathbf{X}} \in R^{n\times p} \mid \sigma(\hat{\mathbf{X}},\mathbf{W},\Delta) \leq \sigma(\mathbf{X},\mathbf{W},\Delta) \text{ for all } \mathbf{X} \in X_{local}(\mathbf{W},\Delta)\} . \tag{1.4}$$

In this paper we focus on the local minimum map for MMDS and global minimum map for full-dimensional MMDS. In particular, we consider the *inverse MMDS* problem, that is, we specify the set of dissimilarity matrices that have the fixed configuration $\mathbf{X}$ as stationary point and the smaller sets for which $\mathbf{X}$ is a local minimum and a global minimum. 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 Δ for a given $\mathbf{X}$. If the set is small then the configuration describes Δ reliably. Conversely, if the set is large, then $\mathbf{X}$ is an unreliable presentation of Δ, since many other Δ have $\mathbf{X}$ as stationary point. We start by describing the maps in more detail.

2. Using Differentiability

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

$$X_{diff}(\mathbf{W},\Delta) \triangleq \{\mathbf{X} \in R^{n\times p} \mid \sigma(\mathbf{X},\mathbf{W},\Delta) \text{ is differentiable at } \mathbf{X}\} . \tag{2.1}$$

De Leeuw (1984) has shown that if the weights and dissimilarities are nonnegative as assumed above, then

$$X_{local}(\mathbf{W},\Delta) \subseteq X_{diff}(\mathbf{W},\Delta) . \tag{2.2}$$

But this means that if

$$X_{stationary}(\mathbf{W},\Delta) \triangleq \{\mathbf{X} \in R^{n \times p} \mid \frac{\partial \sigma(\mathbf{X},\mathbf{W},\Delta)}{\partial \mathbf{X}} = \mathbf{0}\} , \tag{2.3}$$

then

$$X_{local}(\mathbf{W},\Delta) \subseteq X_{stationary}(\mathbf{W},\Delta) . \tag{2.4}$$

Most MMDS algorithms use gradient or subgradient type methods to find a configuration in $X_{stationary}(\mathbf{W},\Delta)$, and then hope it will also be in $X_{local}(\mathbf{W},\Delta)$. This is not necessarily true, of course. We can have vanishing partials in saddle points as well (De Leeuw (1993) shows that STRESS has no local maxima). Actually, we have to be a bit more precise here. The MMDS algorithms look for configurations with

$$\left\| \frac{\partial \sigma(\mathbf{X},\mathbf{W},\Delta)}{\partial \mathbf{X}} \right\| < \varepsilon \tag{2.5}$$

for some small $\varepsilon > 0$. 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 possibility makes it necessary 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 nonnegative definite, and where it is positive definite. We write them as

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

$$X_{pos-hes}(\mathbf{W},\Delta) \triangleq \{\mathbf{X} \in R^{n \times p} \mid \frac{\partial^2 \sigma(\mathbf{X},\mathbf{W},\Delta)}{\partial \mathbf{X}^2} \gtrdot 0\} .$$

It follows that

$$X_{pos-hes}(\mathbf{W},\Delta) \cap X_{stationary}(\mathbf{W},\Delta) \subseteq X_{local}(\mathbf{W},\Delta)$$
$$\subseteq X_{nne-hes}(\mathbf{W},\Delta) \cap X_{stationary}(\mathbf{W},\Delta) . \tag{2.8}$$

This formalization just says that a necessary condition for a configuration to be a local minimum is that the partials vanish and the Hessian is nonnegative definite; a sufficient condition is that the partials vanish and the Hessian is positive definite. Let

$$X_{l\text{-}local}(\mathbf{W},\Delta) \triangleq X_{pos\text{-}hes}(\mathbf{W},\Delta) \cap X_{stationary}(\mathbf{W},\Delta)\,; \tag{2.9}$$

$$X_{u\text{-}local}(\mathbf{W},\Delta) \triangleq X_{nne\text{-}hes}(\mathbf{W},\Delta) \cap X_{stationary}(\mathbf{W},\Delta)\,. \tag{2.10}$$

Then, instead of studying X_{local} directly, we can study $X_{stationary}$ or $X_{l\text{-}local}$ and $X_{u\text{-}local}$. These maps are far from simple. De Leeuw (1993) has shown that STRESS has local minima, sharp ridges, and other irregularities. There seems to be no obvious relationship among the different local minima, and there are no systematic results on the number of local minima. 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

To understand the mappings $X_{stationary}$, $X_{l\text{-}local}$ and $X_{u\text{-}local}$ a bit better, we look at their inverses. Thus, instead 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. One reason is that the inverse maps turn out to be comparatively simple. And by studying them in detail, we learn much about the maps themselves. There is a useful analogy. 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. MMDS is quite close to an eigenvalue problem in various aspects (De Leeuw 1977), but versions of MMDS that use S-STRESS or STRAIN are even more like eigenvalue problems. The inverse MMDS problem for S-STRESS is discussed in Groenen, De Leeuw, and Mathar (1996).

For the present, we restrict ourselves to configurations $\mathbf{X}$ which have $d_{ij} > 0$ for all $i \neq j$. Since we are interested in local minima, this restriction causes no real loss of generality (De Leeuw 1984). The inverse of $X_{stationary}$, for instance, is defined as

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

Inverses for the other maps are defined in the same way, but we analyze only the partial-map in this section. To do so efficiently, we also define

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

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

Because $\partial\sigma(\mathbf{X},\mathbf{W},\Delta)/\partial\mathbf{X}$ is used often in this paper, we give a convenient representation here, using the notation familiar from such earlier papers as De Leeuw and Heiser (1980), De Leeuw (1988), and Groenen et al. (1996). Let $\mathbf{X}_s$ be column s ($s = 1, \ldots, p$) of $\mathbf{X}$. Then, the squared Euclidean distance can be written as

$$d_{ij}^2(\mathbf{X}) = \sum_{s=1}^{p} (x_{is} - x_{js})^2 = \sum_{s=1}^{p} \mathbf{x}'_s(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)'\,\mathbf{x}_s$$

$$= \text{tr}\,\mathbf{X}'\mathbf{A}_{ij}\mathbf{X}\,, \tag{3.3}$$

where

$$\mathbf{A}_{ij} \triangleq (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)'\,, \tag{3.4}$$

and with e_i the unit column vectors of R^n. Hence, the partials of the squared Euclidean distance are

$$\frac{\partial\, d_{ij}^2(\mathbf{X})}{\partial\mathbf{X}} = \frac{\partial\,\text{tr}\,\mathbf{X}'\mathbf{A}_{ij}\mathbf{X}}{\partial\mathbf{X}} = 2\mathbf{A}_{ij}\mathbf{X}\,. \tag{3.5}$$

Using both the standard rules for differentiation and (3.5), the partials of STRESS for differentiable $\mathbf{X}$ are

$$\frac{\partial\,\sigma(\mathbf{X},\mathbf{W},\Delta)}{\partial\mathbf{X}} = \sum_{i=1}^{n}\sum_{j=1}^{n} 2w_{ij}\left[1 - \frac{\delta_{ij}}{d_{ij}(\mathbf{X})}\right]\mathbf{A}_{ij}\mathbf{X}\,. \tag{3.6}$$

Theorem 3.1 (Inverse).

$$X^{+}_{stationary}(\mathbf{X},\mathbf{W}) = \{\Delta \in H^{n\times n} \mid \delta_{ij} = d_{ij}\left[1 - \frac{t_{ij}}{w_{ij}}\right]\}\,, \tag{3.7}$$

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

Proof. The stationary equations are obtained by setting (3.6) equal to zero, i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\mathbf{A}_{ij}\mathbf{X} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,\frac{\delta_{ij}}{d_{ij}}\,\mathbf{A}_{ij}\mathbf{X}\,. \tag{3.8}$$

We have to solve (3.8) for Δ for given $\mathbf{X}$ and $\mathbf{W}$. We make the transformation indicated in Theorem 3.1, i.e., we define

$$t_{ij} \triangleq w_{ij} - w_{ij} \frac{\delta_{ij}}{d_{ij}}, \tag{3.9}$$

and we solve for t_{ij}. Equation (3.8) transforms to

$$\sum_{i=1}^{n} \sum_{j=1}^{n} t_{ij} \mathbf{A}_{ij} \mathbf{X} = \mathbf{0}. \tag{3.10}$$

Note that the $\mathbf{A}_{ij}$ are a basis for the symmetric, doubly centered (SDC) matrices of order n, so that any SDC matrix can be expressed as $\Sigma_{i=1}^{n} \Sigma_{j=i+1}^{n} s_{ij} \mathbf{A}_{ij}$. Thus (3.10) is solved if we find all SDC matrices $\mathbf{T}$ such that $\mathbf{TX} = \mathbf{0}$. But that means $\mathbf{T} = \mathbf{KMK}'$, with $\mathbf{M}$ an arbitrary symmetric matrix. Thus there are $\frac{1}{2}(n-r)(n-r-1)$ independent solutions in all.

By assumption, we must have $\delta_{ij} \geq 0$, which implies $t_{ij} \leq w_{ij}$. ■

A brief comment is in order here. The t_{ij} are defined by (3.9) only for $i \neq j$, because t_{ii} in (3.10) is multiplied by the zero matrix $\mathbf{A}_{ii}$. Thus, we can define the t_{ii} completely arbitrarily, without undercutting the validity of (3.10). Therefore, we simply choose them in such a way that $\mathbf{T}$ is SDC.

To facilitate comparison with such other basic MMDS papers as De Leeuw and Heiser (1980, 1982), and De Leeuw (1988), we define

$$\mathbf{V} \triangleq \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \mathbf{A}_{ij}; \tag{3.11}$$

$$\mathbf{B} \triangleq \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{\delta_{ij}}{d_{ij}} \mathbf{A}_{ij}. \tag{3.12}$$

Equation (3.10) translates into $\mathbf{T} = \mathbf{V} - \mathbf{B} = \mathbf{KMK}'$. The Inverse Theorem says that Δ can be reconstructed from $\mathbf{B} = \mathbf{V} - \mathbf{KMK}'$ using the definition in (3.12).

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

Proof. The set $\{\mathbf{M} \mid \mathbf{M}$ real symmetric of order $n-r-1\}$ is a subspace of $R^{(n-r-1)\times(n-r-1)}$. Because the mapping $\mathbf{M} \to \mathbf{KMK}'$ is linear, the image is a subspace of $R^{n\times n}$. Imposing the constraints $t_{ij} \leq w_{ij}$ means taking the intersection of this subspace with several half-spaces, which yields a set that is closed, polyhedral, and convex. Obviously $\mathbf{D}(\mathbf{X}) \in X^{+}_{stationary}(\mathbf{X},\mathbf{W})$. Only boundedness is nontrivial. The set of all $\mathbf{T}$ is unbounded if and only is it contains a ray (Rockafellar 1970; Theorem 8.4, p.64), which is a set of the form $\lambda\mathbf{T}_0$ for some $\mathbf{T}_0 \neq \mathbf{0}$ and $\lambda > 0$. Thus, we have to show that $X^{+}_{stationary}(\mathbf{X},\mathbf{W})$ cannot contain a ray. From (3.10) we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij} d_{ij}^2(\mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij} \operatorname{tr} \mathbf{X}'\mathbf{A}_{ij}\mathbf{X}$$

$$= \operatorname{tr} \mathbf{X}' \left[\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij}\mathbf{A}_{ij}\mathbf{X}\right] = \operatorname{tr} \mathbf{X}'\mathbf{0} = 0 . \tag{3.13}$$

Because distances are nonnegative, (3.13) implies that not all t_{ij} can have the same sign. There exists at least one pair with $t_{ij} > 0$ and $t_{kl} < 0$. For this pair there exists a $\lambda > 0$ such that $\lambda t_{ij} > w_{ij}$, and thus the set of matrices **T** cannot be unbounded. ■

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

Proof. We have $\delta_{ij1} \leq \delta_{ij2}$ if and only if $t_{ij1} \leq t_{ij2}$. But, from (3.13),

$$\sum_{i=1}^{n}\sum_{j=1}^{n} (t_{ij1} - t_{ij2}) d_{ij}^2(\mathbf{X}) = 0 , \tag{3.14}$$

which is impossible unless $\mathbf{T}_1 = \mathbf{T}_2$. ■

Corollary 3.4 (Only).

$$X^+_{stationary}(\mathbf{X}) = \{\mathbf{W} \in H^{n\times n}, \Delta \in H^{n\times n} \mid \Delta \in X^+_{stationary}(\mathbf{X},\mathbf{W})\} . \tag{3.15}$$

Proof. Directly from the representation in Theorem 3.1. ■

From the last corollary we can choose **W** arbitrarily in $H^{n\times n}$, 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 Δ_1 and Δ_2 are two different elements of $X^+_{stationary}$, then we can measure their distance by using

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(\delta_{ij1} - \delta_{ij2})^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 (t_{ij1} - t_{ij2})^2 . \tag{3.16}$$

The right side of (3.16) is obtained by substituting for δ_{ij} the definition in (3.7), i.e., $d_{ij}(1 - t_{ij}/w_{ij})$. In particular, the squared distance between Δ and **D(X)**, which of course is simply STRESS, is equal to

$$\sigma(\mathbf{X},\mathbf{W},\Delta) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2\, t_{ij}^2 . \tag{3.17}$$

Now suppose we have m configurations $\mathbf{X}_1, \ldots, \mathbf{X}_m$. We can seek for the set of dissimilarity matrices for which all $\mathbf{X}_j$ are stationary points. It is not necessary that all $\mathbf{X}_j$ have the same rank. Each of the configurations $\mathbf{X}_j$ defines $\mathbf{T}_j$, where $\mathbf{T}_j$ spans an affine space of dimension $\frac{1}{2}(n - r_j)(n - r_j - 1)$ with r_j the rank of $\mathbf{X}_j$ (see the Inverse Theorem 3.1). Any $n \times n$ SDC matrix spans a space of $\frac{1}{2}n(n - 1)$ dimensions. If these spaces of $\mathbf{T}_j$ are "in general position" (that is, linear independent and not parallel), they have an intersection if $\frac{1}{2}\Sigma_{j=1}^{m}(n - r_j)(n - r_j - 1) \geq \frac{1}{2}(m - 1)n(n - 1)$. If all r_j are equal to p this works out to

$$M \leq \frac{n(n-1)}{n(n-1)-(n-p)(n-p-1)}\,. \tag{3.18}$$

It is tempting to speculate that (3.18) is an upper bound on the number of stationary points of STRESS, but the reasoning here is difficult to make rigorous.

4. Computing the Inverse Map

We now go into more detail in describing the convex polyhedron of Δ (see the Bounded Corollary) defined in the Inverse Theorem 3.1. From the computational point of view, it is convenient to use a basis $\{\mathbf{P}_l\}$ for the symmetric matrices of order $n - r - 1$. For example, let $\mathbf{X}$ be a rank $r = 2$ matrix for $n = 5$ objects. The Inverse Theorem states that $\mathbf{T}$ is of at most rank 2, so that $\mathbf{P}_l$ is a 2×2 matrix. In this example a basis for the symmetric matrices of order 2 consists of

$$\mathbf{P}_1 = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\ \mathbf{P}_2 = \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix},\ \text{and } \mathbf{P}_3 = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$

From the basis $\{\mathbf{P}_l\}$ a basis $\{\mathbf{Q}_l\}$ for $\mathbf{T}$ can be defined as

$$\mathbf{Q}_l \triangleq \mathbf{K}\mathbf{P}_l\mathbf{K}'\,. \tag{4.1}$$

Then every $\mathbf{T}$ can be written as a weighted sum of the basis $\{\mathbf{Q}_l\}$, i.e.,

$$\mathbf{T} = \sum_{l=1}^{L} \theta_l \mathbf{Q}_l\,. \tag{4.2}$$

But not every θ yields an admissible $\mathbf{T}$. If we limit ourselves to the case $w_{ij} = 1$ for all $i \neq j$, then the Inverse Theorem states that $t_{ij} \leq 1$, so that we must have

$$\sum_{l=1}^{L} \theta_l q_{ijl} \leq 1 \tag{4.3}$$

for all $i < j$, which are $N \triangleq \frac{1}{2}n(n - 1)$ linear inequalities in

$L = \frac{1}{2}(n-r)(n-r-1)$ unknowns, where q_{ijl} contains the elements of $\mathbf{Q}$. Obviously, these linear inequalities describe the bounded convex polyhedron of the Bounded Corollary.

Bounded convex polyhedra can be described in terms of their vertices; compare Goldsman and Tucker (1956) and Tschernikow (1971, p.83-84). We find the vertices of the polyhedron by an enumerative procedure which looks at all subsystems of L rows of (4.3). If the complete system is written as $\mathbf{Q}\theta \geq -\mathbf{u}$, then we can write a subsystem as

$$\begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \theta \geq - \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \tag{4.4}$$

with $\mathbf{Q}_1$ of order L. We then check if $\mathbf{Q}_1$ is singular. If it is, we go to the next subsystem. If it is nonsingular, we compute $\tilde{\theta} = -\mathbf{Q}_1^{-1}\mathbf{u}_1$. If $\mathbf{Q}_2\tilde{\theta} \geq -\mathbf{u}_2$ we add $\tilde{\theta}$ to our list of vertices. If not, we go to the next subsystem. This procedure 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 (Dantzig 1963). In the example below (and those in the Appendix), we simply use brute force, and investigate all subsets.

We start with a really simple example. Call it the Square Example. Consider the configuration

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

We now want to find all dissimilarity matrices Δ for which $\mathbf{X}$ gives a stationary value of STRESS. For $\mathbf{K}$ we find

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

with $-1 \le \theta \le +1$. Using $\delta_{ij} = d_{ij}(1 - t_{ij}/w_{ij})$ from the Inverse Theorem 3.1 and $w_{ij} = 1$, the two extreme points are

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

By defining $\lambda = (1 - \theta)/2$, we can say that any convex combination $\Delta(\lambda) = \lambda\Delta_1 + (1 - \lambda)\Delta_2$ of these vertices has the four points arranged on a ''pseudosquare,'' which gives the solution of our problem. The length of each side is equal to $2(1 - \lambda)$ and the length of each diagonal to $2\lambda\sqrt{2}$. It follows directly from this interpretation that for $\Delta(\lambda)$ to be embeddable in Euclidean space we need to have $\lambda \le \frac{1}{2}$, while $\Delta(\lambda)$ satisfies all triangle inequalities for $\lambda \le 2/(2 + \sqrt{2}) \approx .586$. The distance matrix $\mathbf{D} = \Delta(\frac{1}{2})$ is exactly in the middle of the edge. For $\lambda = 1/(1 + \sqrt{2}) \approx .414$, we have the matrix with all six dissimilarities equal, which is the distance matrix of a regular simplex in three dimensions. For the squared distances between the vertices and their centroid $\mathbf{D}$, we find

	$\mathbf{D}$	Δ_1	Δ_2
$\mathbf{D}$	0	8	8
Δ_1	8	0	32
Δ_2	8	32	0

Thus the STRESS of both vertices is 8.

5. Improved Approximation

We know that $X^{+}_{stationary}(\mathbf{X},\mathbf{W})$ is a compact convex set. It is clear from Equation (2.8) that $X_{u\text{-}local}(\mathbf{X},\mathbf{W})$ is a more precise approximation of $X_{local}(\mathbf{X},\mathbf{W})$, 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, discussed earlier in De Leeuw (1988) and Groenen et al. (1996). We start with equation (3.5):

$$\frac{\partial \sigma}{\partial \mathbf{x}_s} = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\mathbf{A}_{ij}\mathbf{x}_s - 2\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{\delta_{ij}}{d_{ij}} \mathbf{A}_{ij}\mathbf{x}_s . \tag{5.1}$$

Let the superscripted δ be the Kronecker symbol, which has nothing to do with the subscripted δ's. Then

$$\frac{\partial^2 \sigma}{\partial \mathbf{x}_s \partial \mathbf{x}_t} = 2\delta^{st} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\mathbf{A}_{ij} - 2\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{\delta_{ij}}{d_{ij}} \mathbf{A}_{ij} \right\}$$

$$+ 2\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{\delta_{ij}}{d_{ij}^3} \mathbf{A}_{ij}\mathbf{x}_s\mathbf{x}'_t\mathbf{A}_{ij} . \tag{5.2}$$

From the definition (3.9) of t_{ij},

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

Make this substitution twice in (5.2). Note that

$$\begin{aligned} \mathbf{A}_{ij}\mathbf{x}_s\mathbf{x}'_t\mathbf{A}_{ij} &= (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)'\mathbf{x}_s\mathbf{x}'_t(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)' \\ &= (x_{is} - x_{js})(x_{it} - x_{jt})(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)' \\ &= (x_{is} - x_{js})(x_{it} - x_{jt})\mathbf{A}_{ij} , \end{aligned} \tag{5.4}$$

which yields

$$\frac{\partial^2 \sigma}{\partial \mathbf{x}_s \partial \mathbf{x}_t} = 2\delta^{st}\mathbf{T} - 2\sum_{i=1}^{n} \sum_{j=1}^{n} t_{ij} \frac{(x_{is} - x_{js})(x_{it} - x_{jt})}{d_{ij}^2} \mathbf{A}_{ij}$$

$$+ 2\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{(x_{is} - x_{js})(x_{it} - x_{jt})}{d_{ij}^2} \mathbf{A}_{ij} . \tag{5.5}$$

It is convenient at this point to define the $np \times np$ supermatrices $\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_L$ with submatrices

$$\mathbf{H}_{0st} \triangleq \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{(x_{is} - x_{js})(x_{it} - x_{jt})}{d_{ij}^2} \mathbf{A}_{ij} ; \tag{5.6}$$

$$\mathbf{H}_{lst} \triangleq \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ijl} \frac{(x_{is} - x_{js})(x_{it} - x_{jt})}{d_{ij}^2} \mathbf{A}_{ij} , \tag{5.7}$$

where $\mathbf{Q}_l$ is given by (4.1). Also

$$\overline{\mathbf{Q}}_l \triangleq \underbrace{\mathbf{Q}_l \oplus \cdots \oplus \mathbf{Q}_l}_{p\,times} \tag{5.8}$$

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

$$\frac{\partial^2 \sigma}{\partial \mathbf{X}^2} = 2\mathbf{H}_0 + 2\sum_{l=1}^{L} \theta_l(\overline{\mathbf{Q}}_l - \mathbf{H}_l)\,. \tag{5.9}$$

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

Proof. We have $\Delta \in X^+_{u\text{-}local}(\mathbf{X},\mathbf{W})$ if and only if Δ is of the form in Inverse Theorem 3.1, with, in addition,

$$\sum_{l=1}^{L} \theta_l(\overline{\mathbf{Q}}_l - \mathbf{H}_l) \geq -\mathbf{H}_0\,. \tag{5.10}$$

But this statement means that $X^+_{u\text{-}local}(\mathbf{X},\mathbf{W})$ is the intersection of the convex set defined by (5.10) and the compact convex set from the Inverse Theorem; i.e., it is a compact convex set. ■

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

In our Square Example, we can still carry out the necessary computations quite easily. We know from the results of De Leeuw (1988, p. 173) that the Hessian has at least $1/2p(p+1)$ eigenvalues equal to zero, corresponding to the rotational and translational invariance of Euclidean distances, and it has at least one eigenvalue equal to n. Recall the meanings of θ and λ in the example. Then the smallest eigenvalue is equal to zero for $-1/2 \leq \theta \leq 1$, and it is negative for $-1 \leq \theta \leq -1/2$, i.e., $3/4 \leq \lambda \leq 1$. Thus, the more precise approximation tells us that for a local minimum we must have $0 \leq \lambda \leq 3/4$, while for any stationary point, it suffices to have $0 \leq \lambda \leq 1$. At $\lambda = 3/4$, the STRESS is 2, and

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

More examples of inverse scaling are given in the Appendix.

6. Full-dimensional Scaling

In MDS we minimize STRESS over $\mathbf{D}$, on the condition that $\mathbf{D} = \mathbf{D}(\mathbf{X})$, i.e., $\mathbf{D}$ are the Euclidean distances between pairs of the n points of a configuration in p dimensions. Now suppose we drop the constraint of p dimensions and merely require that the $\mathbf{D}$ are Euclidean distances between points in any configuration. This framework defines *metric full-dimensional scaling*, or MFDS. There is no need to emphasize the fact that MFDS is metric, because nonmetric full-dimensional scaling does not make sense, given that any dissimilarity matrix can be fitted perfectly in $n-2$ dimensions nonmetrically (Lingoes 1971). The most interesting result on MFDS is that all local minima are global. This result is due to De Leeuw (1993), but because the proof is difficult to find and simple to reproduce, we give it here for completeness.

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

Proof. Let $\mathbf{C} = \mathbf{XX}'$, where $\mathbf{X}$ is of rank $n-1$ and column centered, so that $\mathbf{C}$ is positive semidefinite and of rank $n-1$ as well. Using (3.3), d_{ij} can be written as

$$\begin{aligned} d_{ij} &= d_{ij}(\mathbf{X}) = (\operatorname{tr}\mathbf{X}'\mathbf{A}_{ij}\mathbf{X})^{1/2} = (\operatorname{tr}\mathbf{A}_{ij}\mathbf{XX}')^{1/2} \\ &= (\operatorname{tr}\mathbf{A}_{ij}\mathbf{C})^{1/2} = (c_{ii} + c_{jj} - 2c_{ij})^{1/2} = d_{ij}(\mathbf{C}). \end{aligned} \tag{6.1}$$

Thus, the distance d_{ij} can be specified either by $d_{ij}(\mathbf{X})$ or by $d_{ij}(\mathbf{C})$ as defined in (6.1). The MFDS problem minimizes

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

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \delta_{ij}^2 + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} d_{ij}^2(\mathbf{C})$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \delta_{ij} d_{ij}(\mathbf{C}) \tag{6.2}$$

over the convex cone of all positive semidefinite matrices **C**. Each of the three terms of (6.2) is convex, which can be seen as follows. The first term is constant, which is convex in **C**. The second term is convex because it is linear in **C** as $d_{ij}^2(\mathbf{C}) = \operatorname{tr} \mathbf{A}_{ij}\mathbf{C}$. Now, the third term uses $d_{ij}(\mathbf{C})$, which is the square root of a linear function of **C**. Therefore, $d_{ij}(\mathbf{C})$ is concave in **C**, so that $-d_{ij}(\mathbf{C})$ is convex in **C**. Thus, the third term is also convex. It follows that $\sigma(\mathbf{C})$ is convex in **C**, and thus the MFDS problem minimizes a convex function over a convex set. All local minima are global. ∎

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

$$\mathbf{x}_{full}(\mathbf{W},\Delta) \triangleq \{\mathbf{X} \in R^{n \times p} \mid \mathbf{X} \text{ solves the MFDS problem } \} ; \tag{6.3}$$

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

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

Proof. If we minimize a differentiable convex function $f(\cdot)$ over a convex cone K, then the necessary and sufficient conditions for a minimum $\hat{\mathbf{x}}$ (Rockafeller, 1970, Theorem 27.4, p. 270-271) are (1) $\hat{\mathbf{x}} \in k$; (2) $-\nabla f(\hat{\mathbf{x}}) \in K^{\circ}$; and (3) $\hat{\mathbf{x}}' \nabla f(\hat{\mathbf{x}}) = 0$; that is, $\hat{\mathbf{x}}$ must be in the convex cone, minus the gradient of $\hat{\mathbf{x}}$ must be in the polar cone, and the $\hat{\mathbf{x}}$ should be orthogonal to its gradient. In our case this requirement means that the necessary and sufficient conditions for the MFDS problem are

$$\mathbf{C} \gtrsim 0 ; \tag{6.5}$$

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

$$\operatorname{tr} \mathbf{C} \{ \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \mathbf{A}_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{w_{ij} \delta_{ij}}{d_{ij}(\mathbf{C})} \mathbf{A}_{ij} \} = 0 . \tag{6.7}$$

Condition (6.5) states that **C** must be a double centered positive semidefinite matrix of at most rank $n - 1$. The polar cone of this set is the set of double centered negative semidefinite matrices of at most rank $n - 1$. Condition (6.6) states that the gradient of a $\sigma(\mathbf{C})$ must be in the set of double centered positive semidefinite matrices of at most rank $n - 1$. Finally, condition (6.7) requires that $\mathbf{C}$ is orthogonal to its gradient. Condition (6.6) implies that $\mathbf{T} \geq 0$ and condition (6.7) that $\mathbf{TX} = 0$. But this situation is the same as $\mathbf{T} = \mathbf{KMK}'$, with $\mathbf{M}$ a positive semidefinite matrix. So again we have the intersection of a convex cone and the compact convex set of Inverse Theorem 3.1. ■

For our Square Example we have $\mathbf{T} \geq 0$ if and only if $\theta \geq 0$. Thus the dissimilarity matrices for which $\mathbf{X}$ solves the MFDS problem are the ones on the line segment between $\mathbf{D}$ and Δ_2.

Corollary 6.3 (Maximum Dimensionality). *If rank* $(\mathbf{X}) = n - 1$ *then* $\Delta = \mathbf{D}(\mathbf{X})$.

Proof. The Inverse Theorem 3.1 states that $\mathbf{K}$ is of rank $n - r - 1$. If rank($\mathbf{X}$) $= n - 1$, then the rank($\mathbf{K}$) $= 0$, so that $\mathbf{T} = \mathbf{0}$. Therefore, $\mathbf{D}(\mathbf{X})$ is the only element in X^+_{full} $(\mathbf{W},\Delta)$. ■

Thus, if a rank($\mathbf{X}$) $= n - 1$ and $\mathbf{X}$ is a minimum MFDS, then we must have $\Delta = \mathbf{D}(\mathbf{X})$. This observation implies that $\sigma(\mathbf{X}) = 0$ for all rank($\mathbf{X}$) $= n - 1$ minima of STRESS. The converse is also true. If in the FMDS problem the minimum has $\sigma(\mathbf{X}) > 0$ then the rank($\mathbf{X}$) $\leq n - 2$.

Appendix

We reported two examples of inverse scaling 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

$$\mathbf{X}_l = \begin{bmatrix} -3 \\ -1 \\ +1 \\ +3 \end{bmatrix},$$

whose null space is spanned by

$$\mathbf{K}_l = \begin{bmatrix} +2 & +1 \\ -2 & -3 \\ -2 & +3 \\ +2 & -1 \end{bmatrix}.$$

We have found seven vertices that produce dissimilarities with $\mathbf{X}_l$ as local minima. They are local minima, because the STRESS is a piecewise quadratic function over $\mathbf{X}$, where the pieces depend only on the order of the coordinates of $\mathbf{X}$. The vertices are summarized in Table 1.

Table 1: The vertices of the polyhedral set that defines dissimilarities $\mathbf{\Delta}$ for which $\mathbf{X}_l$ is a local minimum.

vertex	$\mathbf{M}$	STRESS	$\mathbf{\Delta}$
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 $\mathbf{\Delta}$ for which $\mathbf{X}_s$ is a stationary point. λ_i is the i-th eigenvalue of the Hessian $\mathbf{H}$.

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

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 full-dimensional scaling solution
a.	four points equally spaced on a line	7	7	1
b.	equilateral 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

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

$$\mathbf{X}_s = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ +\frac{1}{2} & -\frac{1}{2} \\ +\frac{1}{2} & +\frac{1}{2} \\ -\frac{1}{2} & +\frac{1}{2} \\ 0 & 0 \end{bmatrix}, \text{ which has null space } \mathbf{K}_s = \begin{bmatrix} +\frac{1}{2} & -\frac{1}{2(5)^{1/2}} \\ -\frac{1}{2} & -\frac{1}{2(5)^{1/2}} \\ +\frac{1}{2} & -\frac{1}{2(5)^{1/2}} \\ -\frac{1}{2} & -\frac{1}{2(5)^{1/2}} \\ 0 & -\frac{2}{2(5)^{1/2}} \end{bmatrix}.$$

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

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

References

DANTZIG, G. B. (1963), *Linear Programming and Extensions*, Princeton, New Jersey: Princeton University Press.

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.

DE LEEUW, J. (1984), ''Differentiability of Kruskal's Stress at a Local Minimum,'' *Psychometrika, 49*, 111-113.

DE LEEUW, J. (1988), ''Convergence of the Majorization Method for Multidimensional Scaling,'' *Journal of Classification, 5*, 163-180.

DE LEEUW, J. (1993), ''Fitting Distances by Least Squares,'' Technical Report, UCLA Statistics Series #130, Los Angeles: UCLA.

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.

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.

GROENEN, P. J. F. (1993), *The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions*, Leiden: DSWO Press.

GROENEN, P. J. F., DE LEEUW, J., and MATHAR, R. (1996), ''Least Squares Multidimensional Scaling with Transformed Distances,'' in *Studies in Classification, Data Analysis, and Knowledge Organization*, Eds., W. Gaul and D. Pfeifer, Berlin: Springer, 177-185.

GROENEN, P. J. F., and HEISER, W. J. (1996), ''The Tunneling Method for Global Optimization in Multidimensional Scaling,'' *Psychometrika, 61*, 529-550.

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.

KRUSKAL, J. B. (1964a), ''Multidimensional Scaling by Optimizing Goodness of Fit to a Non-metric Hypothesis,'' *Psychometrika, 29*, 1-27.

KRUSKAL, J. B. (1964b), ''Nonmetric Multidimensional Scaling: A Numerical Method,'' *Psychometrika, 29*, 115-129.

GOLDSMAN, A. J., and TUCKER, A. W. (1956), ''Polyhedral Convex Cones,'' in *Linear Inequalities and Related Systems,* Volume 38 of *Annals of Mathematics Studies*, Eds., H.W. Kuhn and A.W. Tucker, Princeton, New Jersey: Princeton University Press, 19-40.

LINGOES, J. C. (1971), ''Some Boundary Conditions for a Monotone Analysis of Symmetric Matrices,'' *Psychometrika, 36*, 195-203

ROCKAFELLER, R. T. (1970), *Convex Analysis*, Princeton, New Jersey: Princeton University Press.

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