RR 86-14

Explicit SMACOF algorithms for individual differences scaling

PROXSCAL Progress Report 2*

Willem J. Heiser
Department of Data Theory
University of Leiden

Ineke Stoop
Department S5
Central Bureau of Statistics

December 1986

* PROXSCAL Progress Reports provide documentation on the design and development of the PROXSCAL Multidimensional Scaling program. They contain preliminary notes, comments, ideas and decisions, without aiming at full explanations.

Explicit SMACOF algorithms for individual differences scaling

Willem J. Heiser

and

Ineke Stoop

Abstract

The SMACOF majorization approach to Multidimensional Scaling in the STRESS framework (De Leeuw and Heiser, 1980) is applied to various well-known individual differences models. The resulting algorithms are described in detail, with particular attention for computational efficiency. Choices that were made in the development of PROXSCAL are indicated.
Special attention is given to the mathematical structure of the REDUCED RANK model. Furthermore a new model is introduced, called the DIAGONAL REDUCED RANK model, which is a special case of both the INDSCAL and the REDUCED RANK model. Parameter estimation for this case requires only one simple additional step. All developments are equally valid for missing data, nonmetric analyses, unfolding, and other a priori or data driven weighting schemes.

1.Introduction

All current models for individual differences scaling of proximities assume that the individual spaces $\mathbf{X}_k$ that account for the *dissimilarities* Δ_k of *data source* k ($k = 1,...,m$) can be described with linear transformations $\mathbf{A}_k$ of a *common space* $\mathbf{Z}$. Each of the n rows of the nxp matrix $\mathbf{Z}$ contains the p coordinates of an *object point* $\mathbf{z}_i$; thus i = 1,...,n, and without risk of ambiguity z_{ia} will denote either the (i,a)'th element of $\mathbf{Z}$ or the a'th element of the p-vector $\mathbf{z}_i$, the representation of object i. The parameter p indicates the *dimensionality* of the model, the optimal choice of which does have great data analytical interest, but, nevertheless, p will be regarded in this report as a fixed, prechosen integer larger than one. Its determination is considered to be an "outermost" problem, possibly attackable by repeated execution and careful evaluation of the computational schemes to be presented.

When each data source is analyzed individually, the goal would be minimizing some Multidimensional Scaling (MDS) loss function for each k. The type of loss function that will be used throughout this report looks like

$$\sigma^2(\mathbf{X}_k) = \sum_{(i,j)\in I(k)} w_{ijk} (\delta_{ijk} - d_{ij}(\mathbf{X}_k))^2 ; \qquad (1)$$

i.e., the weighted sum of squared residuals, defined here separately for each source. Function (1), called *(raw) STRESS* by Kruskal (1964a,b), compares the *Euclidean distance* $d_{ij}(\mathbf{X}_k)$ between the ith and jth row of $\mathbf{X}_k$ with the (i,j)th element of the data matrix Δ_k, and aggregates the numerical deviations across all pairs i,j from some index set specific for source k, $I(k)$ (all ordered pairs, or some subset of these). The weights w_{ijk} can be *a priori* given, they might have an empirical origin (e.g., a measure of the reliability of δ_{ijk}), or they might carry information on certain normalization factors. The latter case arises when *sublists* of the dissimilarities (e.g., the elements of Δ_k partitioned row-wise, or column-wise, or in groups of three) are allowed to be transformed, or, put in slightly more general terms, when sublists are allowed to be *optimally scaled* within a given class of *admissible quantifications* (cf. de Leeuw and Heiser, 1977). Thus, even though attention will be restricted to the case of fixed δ_{ijk} and w_{ijk}, the results of this report can be easily generalized to these so-called *conditional nonmetric* cases (including the one of optimally scaling *all* elements of Δ_k that are present in $I(k)$ at once). Separate analyses of the m data sources - minimizing STRESS m times - would leave us with the problem of comparing m configurations whithout further reference to the dissimilarities from which they were derived.

The aim of individual differences modelling is to relate the m configurations $\mathbf{X}_k$ immediately to each other by setting up a *model for joint analysis*, in which these individual spaces are restricted to be only moderately different in a specified sense, in particular to be transformations of some common space $\mathbf{Z}$. The models to be discussed all have the form

$$\mathbf{X}_k = \mathbf{Z}\mathbf{A}_k , \qquad (2)$$

with $\mathbf{A}_k$ a pxp matrix. The three most important cases are:

$\mathbf{A}_k$ is any full rank matrix	(FULL model);
$\mathbf{A}_k$ is of rank r with $r<p$	(REDUCED rank model);
$\mathbf{A}_k$ is any diagonal matrix	(DIAGONAL model).

The FULL case is better known under the name IDIOSCAL, and was first proposed by Carroll and Chang (1970); the FULL and REDUCED case together have also been called the General Euclidean Model (GEM), cf. Bloxom (1978) and Young (1984) for more elaborate reviews; the

diagonal case is usually called INDSCAL, it was first considered as a model by Horan (1969), while Carroll and Chang (1970) provided the first (and in fact more general) method for fitting all its parameters. Carroll and Chang's methods are generalizations of *classical scaling*, based on the STRAIN loss function (cf. Carroll and Wish, 1974). Young's methods are generalizations of *squared distance scaling*, based on the SSTRESS loss function (Takane et al., 1977). The possibility to fit these individual differences models in the STRESS framework was outlined in Bloxom (1978), de Leeuw (1980), and de Leeuw and Heiser (1980), but their methods were not presented in great detail. The purpose of this report is to offer more explicit algorithms for minimizing generalized STRESS under restrictions (2), with $\mathbf{Z}$ free, as well as with $\mathbf{Z}$ further restricted in various ways.

Generalized STRESS is simply obtained from (1) by averaging over k:

$$\sigma^2(\mathbf{X}_1,\ldots,\mathbf{X}_m) = 1/m \sum_k \sum_{(i,j)\in I(k)} w_{ijk} (\delta_{ijk} - d_{ij}(\mathbf{X}_k))^2 . \tag{3}$$

In section 2 the rationale of the SMACOF algorithm (de Leeuw and Heiser, 1980) will be explained; in particular, it will be shown that (3) can be minimized under restrictions (2) by repeatedly (and in practice only partly) minimizing another function of simpler form, called the *majorizing function*:

$$\mu^2(\mathbf{Z};\mathbf{A}_1,\ldots,\mathbf{A}_m) = 1/m \sum_k \{ c_k + \text{tr } \mathbf{A}_k'\mathbf{Z}'\mathbf{V}_k\mathbf{Z}\mathbf{A}_k - 2 \text{ tr } \mathbf{A}_k'\mathbf{Z}'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \} , \tag{4}$$

where ${}^0\mathbf{X}_k$ is the current estimate of the configuration for source k; the constants c_k, $\mathbf{V}_k$, and the matrix operation $\mathbf{B}_k(.)$ will be defined shortly. At this point it is important to note that $\mu^2(\mathbf{Z};\mathbf{A}_1,\ldots,\mathbf{A}_m)$ is *quadratic* in natural subsets of its unknowns, for this makes the subminimization problems relatively straight-forward indeed. We can minimize the majorizing function for each parameter subset separately. Section 3 will discuss the solution for $\mathbf{Z}$ with $\mathbf{A}_1,\ldots,\mathbf{A}_m$ fixed at their current values, section 4 develops a variety of solutions for $\mathbf{A}_k$ keeping $\mathbf{Z}$ and the other individual transformations fixed, and section 5 splits the problem of finding the common space coordinates further down so that it becomes clear how to handle additional restrictions.

2. The SMACOF majorization approach.

First consider the basic stress component (1) for any one $\mathbf{X}_k$ in matrix notation. We may assume $w_{ijk} = 0$ whenever $(i,j) \notin I(k)$, so that the summations range over all pairs i,j. The constant c_k is defined as the weighted sum of squared dissimilarities:

$$c_k = \sum_{i,j} w_{ijk}\delta^2_{ijk} . \tag{5}$$

For the weighted sum of squared distances we find:

$$\sum_{i,j} w_{ijk}d^2_{ij}(\mathbf{X}_k) = \text{tr } \mathbf{X}_k'\mathbf{V}_k\mathbf{X}_k , \tag{6}$$

where the matrix $\mathbf{V}_k$ is defined as $\mathbf{V}_k = {}^o\mathbf{V}_k + {}^d\mathbf{V}_k$, with ${}^o\mathbf{V}_k$ having off-diagonal elements equal to $-(w_{ijk} + w_{jik})$ and diagonal elements equal to zero, and with ${}^d\mathbf{V}_k$ a diagonal matrix containing the row sums of ${}^o\mathbf{V}_k$. The weighted cross product term can be written as:

$$\sum_{i,j} w_{ijk}\delta_{ijk}d_{ij}(\mathbf{X}_k) = \text{tr } \mathbf{X}_k'\mathbf{B}_k(\mathbf{X}_k)\mathbf{X}_k , \tag{7}$$

where the matrix $\mathbf{B}_k$ is again built up from an off-diagonal and a diagonal part, $\mathbf{B}_k = {}^o\mathbf{B}_k + {}^d\mathbf{B}_k$ with ${}^o\mathbf{B}_k$ having elements

$${}^ob_{ijk} = -(w_{ijk}\delta_{ijk} + w_{jik}\delta_{jik}) / d_{ij}(\mathbf{X}_k) \qquad \text{if} \qquad d_{ij}(\mathbf{X}_k) \neq 0 , \tag{8a}$$

$${}^ob_{ijk} = 0 \qquad \text{if} \qquad d_{ij}(\mathbf{X}_k) = 0 , \tag{8b}$$

and the non-zero elements of ${}^d\mathbf{B}_k$ equal to the row sums of ${}^o\mathbf{B}_k$. The notation $\mathbf{B}_k(\mathbf{X}_k)$ is used to emphasize the dependence on $\mathbf{X}_k$. Thus we obtain

$$\sigma^2(\mathbf{X}_k) = c_k + \text{tr } \mathbf{X}_k'\mathbf{V}_k\mathbf{X}_k - 2 \text{ tr } \mathbf{X}_k'\mathbf{B}_k(\mathbf{X}_k)\mathbf{X}_k . \tag{9}$$

The majorization approach is based on locally approximating $\sigma^2(\mathbf{X}_k)$ from above by a simpler function that has the same function value at the location of the current best estimate ${}^0\mathbf{X}_k$. The function $\mu^2(\mathbf{X}_k, {}^0\mathbf{X}_k)$ majorizes $\sigma^2(\mathbf{X}_k)$ if we have $\sigma^2(\mathbf{X}_k) \leq \mu^2(\mathbf{X}_k, {}^0\mathbf{X}_k)$ for all $\mathbf{X}_k$ and ${}^0\mathbf{X}_k$, and it has the same function value if $\sigma^2({}^0\mathbf{X}_k) = \mu^2({}^0\mathbf{X}_k, {}^0\mathbf{X}_k)$ for all (admissible) choices of ${}^0\mathbf{X}_k$. The update ${}^+\mathbf{X}_k$ of ${}^0\mathbf{X}_k$ is defined as the configuration that minimizes (or, in most practical cases, at least decreases the value of) $\mu^2(\mathbf{X}_k, {}^0\mathbf{X}_k)$. This way we always obtain the chain

$$\sigma^2({}^+\mathbf{X}_k) \leq \mu^2({}^+\mathbf{X}_k, {}^0\mathbf{X}_k) \leq \mu^2({}^0\mathbf{X}_k, {}^0\mathbf{X}_k) = \sigma^2({}^0\mathbf{X}_k) , \tag{10}$$

so that convergence is guaranteed.

The SMACOF majorizing function (the acronym means: Scaling by MAjorizing a COmplicated Function) is derived from the Cauchy-Schwarz inequality in the following form:

$$\| \mathbf{x}_{i,k} - \mathbf{x}_{j,k} \| \, \| {}^0\mathbf{x}_{i,k} - {}^0\mathbf{x}_{j,k} \| \geq (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})'({}^0\mathbf{x}_{i,k} - {}^0\mathbf{x}_{j,k}) \, , \tag{11}$$

where the notation i,k is used to denote the i'th point of source k. If we write $d_{ij}(\mathbf{X}_k)$ and $d_{ij}({}^0\mathbf{X}_k)$ for the lengths of the difference vectors on the left-hand side of (11), we obtain

$$d_{ij}(\mathbf{X}_k) \geq (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})'({}^0\mathbf{x}_{i,k} - {}^0\mathbf{x}_{j,k}) \, / \, d_{ij}({}^0\mathbf{X}_k) \, , \tag{12a}$$

$$w_{ijk}\delta_{ijk}d_{ij}(\mathbf{X}_k) \geq w_{ijk}\delta_{ijk} \, (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})'({}^0\mathbf{x}_{i,k} - {}^0\mathbf{x}_{j,k}) \, / \, d_{ij}({}^0\mathbf{X}_k) \, , \tag{12b}$$

$$\{w_{ijk}\delta_{ijk} \, / \, d_{ij}(\mathbf{X}_k)\} \, d^2_{ij}(\mathbf{X}_k) \geq \{w_{ijk}\delta_{ijk} \, / \, d_{ij}({}^0\mathbf{X}_k)\} \, (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})'({}^0\mathbf{x}_{i,k} - {}^0\mathbf{x}_{j,k}) \tag{12c}$$

In (12b) we multiply both sides with the *nonnegative* quantity $w_{ijk}\delta_{ijk}$, so that the inequality indeed remains valid. Similarly, we have to except for the case of vanishing distances, which is the reason for definition (8b): we can still sum (12c) over *all* i and j this way, obtaining after rearrangements and simplifications:

$$\text{tr } \mathbf{X}_k'\mathbf{B}(\mathbf{X}_k)\mathbf{X}_k \geq \text{tr } \mathbf{X}_k'\mathbf{B}({}^0\mathbf{X}_k){}^0\mathbf{X}_k \, . \tag{13}$$

This inequality forms the basis of the SMACOF majorization approach; from (13) it follows that

$$\mu^2(\mathbf{X}_k, {}^0\mathbf{X}_k) = c_k + \text{tr } \mathbf{X}_k'\mathbf{V}_k\mathbf{X}_k - 2 \text{ tr } \mathbf{X}_k'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \tag{14}$$

indeed majorizes STRESS (9), and that (14) is simpler because it is quadratic in $\mathbf{X}_k$. It is sometimes useful to reexpress (14) as

$$\mu^2(\mathbf{X}_k, {}^0\mathbf{X}_k) = c_k + \eta^2(\mathbf{X}_k, {}^0\mathbf{X}_k) - \text{tr } {}^0\mathbb{X}_k'\mathbf{V}_k{}^0\mathbb{X}_k \, , \tag{15}$$

with

$$\eta^2(\mathbf{X}_k, {}^0\mathbf{X}_k) = \text{tr } ({}^0\mathbb{X}_k - \mathbf{X}_k)'\mathbf{V}_k({}^0\mathbb{X}_k - \mathbf{X}_k) \, , \tag{16a}$$

$$\mathbf{V}_k{}^0\mathbb{X}_k = \mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \, , \tag{16b}$$

because (15) shows that minimizing $\mu^2(\mathbf{X}_k, {}^0\mathbf{X}_k)$ for some fixed ${}^0\mathbf{X}_k$ is equivalent to minimizing $\eta^2(\mathbf{X}_k, {}^0\mathbf{X}_k)$, where ${}^0\mathbb{X}_k$ - called the *Guttman transform* of ${}^0\mathbf{X}_k$ - is the solution of the linear system (16b). Now if $\mathbf{X}_k$ is *unconstrained*, it follows from (16a) that the Guttman transform is all what we need; it lets $\eta^2(\mathbf{X}_k, {}^0\mathbf{X}_k)$ vanish. However, if $\mathbf{X}_k$ actually *is* constrained we frequently don't need to solve the quite large (order n) system of equations (16b) first and next minimize (16a) - called the *metric projection* problem (de Leeuw and

Heiser, 1980), as it involves finding the feasible configuration of minimum distance, in the metric $\mathbf{V}_k$, from the Guttman transform. This will become evident in the next sections.

In the majorization algorithm the current configuration is always the update of the previous iteration, so from now on reference to ${}^0\mathbf{X}_k$ in the argument list of the majorizing function will be suppressed. Generalized STRESS, being the average of the (nonnegative) STRESS components (9), is majorized by the average of the functions (14):

$$\mu^2(\mathbf{X}_1,...,\mathbf{X}_m) = 1/m \sum_k \{ c_k + \text{tr } \mathbf{X}_k'\mathbf{V}_k\mathbf{X}_k - 2 \text{ tr } \mathbf{X}_k'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \} . \tag{17}$$

If the general form of the individual differences models, $\mathbf{X}_k = \mathbf{Z}\mathbf{A}_k$, is inserted in this average, we obtain (4), the function presented in the introduction. We now turn to the problem of finding local improvements for the common space $\mathbf{Z}$.

3. Simultaneous solution for the common space coordinates.

The case of all individual spaces *identical*, i.e. $\mathbf{X}_1 = ... = \mathbf{X}_m = \mathbf{Z}$, though not mentioned earlier as a model, is still of some interest, because it is not only conceptually but also computationally quite distinct from the other cases. It allows a great simplification. Inserted in (17) we get:

$$\begin{aligned}\mu^2(\mathbf{Z}) &= 1/m \sum_k \{ c_k + \text{tr } \mathbf{Z}'\mathbf{V}_k\mathbf{Z} - 2 \text{ tr } \mathbf{Z}'\mathbf{B}_k({}^0\mathbf{Z}){}^0\mathbf{Z} \} \\ &= c_* + \text{tr } \mathbf{Z}'\{1/m \sum_k \mathbf{V}_k\}\ \mathbf{Z} - 2 \text{ tr } \mathbf{Z}'\{1/m \sum_k \mathbf{B}_k({}^0\mathbf{Z})\}\ {}^0\mathbf{Z} ,\end{aligned} \tag{18}$$

with c_* the mean of the individual constants c_k. Thus the stationary equations for a new ${}^+\mathbf{Z}$ are

$$\{1/m \sum_k \mathbf{V}_k\}\ {}^+\mathbf{Z} = \{1/m \sum_k \mathbf{B}_k({}^0\mathbf{Z})\}\ {}^0\mathbf{Z} . \tag{19}$$

So we may use the mean $\mathbf{B}$-matrix with respect to the old distances (the numerator of (8a) remains constant), and have to solve only one system of equations, with the mean weight matrix as coefficients. This procedure is used in the programs SMACOF-1 (de Leeuw and Heiser, 1977), SMACOF-1A (Stoop et al., 1981), SMACOF-1B (Stoop and de Leeuw, 1982), and in PROXSCAL under the option MODEL=IDENTICAL, or NSPA = 1.

Now suppose $\mathbf{A}_k \neq \mathbf{I}$; then the majorizing function (4) can be rearranged as

$$\mu^2(\mathbf{Z},\mathbf{A}_1,...,\mathbf{A}_m) = c_* + \text{tr } \mathbf{Z}'\{1/m \sum_k \mathbf{V}_k\mathbf{Z}\mathbf{A}_k\mathbf{A}_k'\} - 2 \text{ tr } \mathbf{Z}'(\tilde{}\mathbf{X}) , \tag{20}$$

with

$$\tilde{}\mathbf{X} = 1/m \sum_k \mathbf{B}_k({}^0\mathbf{X}_k)^0\mathbf{X}_k\mathbf{A}_k' . \tag{21}$$

So the stationary equations that the update ${}^+\mathbf{Z}$ must satisfy become

$$1/m \sum_k \mathbf{V}_k({}^+\mathbf{Z})\mathbf{A}_k\mathbf{A}_k' = \tilde{}\mathbf{X} . \tag{22}$$

Unravelling the common space coordinates dimension after dimension (and $\tilde{}\mathbf{X}$ accordingly) shows that (22) involves solving a linear system of np equations, with a coefficient matrix that is blockwise structured as

$$\begin{array}{llll} 1/m \sum_k {}^kc_{11}\mathbf{V}_k & 1/m \sum_k {}^kc_{12}\mathbf{V}_k & \ldots & 1/m \sum_k {}^kc_{1p}\mathbf{V}_k \\ 1/m \sum_k {}^kc_{21}\mathbf{V}_k & 1/m \sum_k {}^kc_{22}\mathbf{V}_k & \ldots & 1/m \sum_k {}^kc_{2p}\mathbf{V}_k \\ \ldots & \ldots & \ldots & \ldots \\ 1/m \sum_k {}^kc_{p1}\mathbf{V}_k & 1/m \sum_k {}^kc_{p2}\mathbf{V}_k & \ldots & 1/m \sum_k {}^kc_{pp}\mathbf{V}_k \end{array} \tag{23}$$

where ${}^kc_{ab}$ denotes the (a,b)th element of the p x p matrix of sums of squares and cross products $\mathbf{C}_k = \mathbf{A}_k\mathbf{A}_k'$. The solution of this system would constitute a locally optimal, simultaneous improvement of all common space coordinates with the current best estimates of $\mathbf{A}_1,\ldots,\mathbf{A}_m$ held fixed.

In general, this is not an attractive practical way to proceed, because n might be "quite large", as in unfolding (where $n = 150$ is not unusual), and because p might be "not very small" in individual differences modelling (e.g., $p = 7$ would have to be covered), in contrast to ordinary MDS - where 2, 3 or 4 dimensions are typically used. Combining the two examples, we would have to face the challenge of solving "very large" systems of 1050 equations *every iteration*, since even if the $\mathbf{V}_k$s remain constant (which might also not be the case in many applications) the coefficients (23) would continually change, due to their dependence on the $\mathbf{A}_k$s. In SMACOF-3 and in the unfolding options of PROXSCAL a basic principle is *not* to work with the complete $\mathbf{V}_k$ matrices, but to take advantage of their structured sparseness; this principle would have to be dropped, or else a very specialized equation solver would have to be developed for the unfolding case.

There are a number of special cases for which the system simplifies. In the DIAGONAL (or INDSCAL) model the off-diagonal elements of $\mathbf{C}_k$ vanish, so that the coefficient matrix (23) becomes block-diagonal and the common space coordinates can be found dimension after dimension. There is no practical problem here. Secondly, suppose all weight matrices are

equal, i.e., $\mathbf{V}_1 = ... = \mathbf{V}_m = \mathbf{V}$. Then we can make, without loss of generality, the simplying assumption, or, put rather differently, then we can choose the identification condition

$$1/m \sum_k \mathbf{C}_k = \mathbf{I} . \tag{24}$$

The reason is the indeterminacy of the model $\mathbf{X}_k = \mathbf{Z}\mathbf{A}_k$. For we can always define (infinitely many) alternate solutions as

$$^*\mathbf{Z} = \mathbf{Z}\mathbf{T}^{-1} , \tag{25a}$$
$$^*\mathbf{A}_k = \mathbf{T}\mathbf{A}_k , \tag{25b}$$

that render an equivalent set of $\mathbf{X}_1,...,\mathbf{X}_m$. Thus we must somehow agree upon how to choose the transformation $\mathbf{T}$ (the final solution can always be readjusted to satisfy some other convention). In PROXSCAL this is done as follows. Suppose $^*\mathbf{A}_1,...,^*\mathbf{A}_m$ are the untransformed estimates of the individual space transformations; then we compute

$$1/m \sum_k {}^*\mathbf{A}_k {}^*\mathbf{A}_k' = \mathbf{T}\mathbf{T}' , \tag{26a}$$
$$^+\mathbf{A}_k = \mathbf{T}^{-1}(^*\mathbf{A}_k) \text{ for } k = 1,...,m . \tag{26b}$$

In particular, $\mathbf{T}$ is chosen as the Cholesky factor of the matrix of sums of squares and cross products in (26a); it can easily be verified that the $^+\mathbf{A}_1,...,^+\mathbf{A}_m$ of (26b) do satisfy (24). In the DIAGONAL case this amounts to rescaling the dimension weights to have unit mean squares; it may be remarked here that in the INDSCAL method and program (Carroll and Chang, 1970) the convention is used to choose $\mathbf{T}$ such that the columns of $\mathbf{Z}$ have unit sums of squares. When the individual transformation matrices satisfy (24) and all weight matrices are equal, the stationary equations (22) simplify to

$$\mathbf{V}(^+\mathbf{Z}) = \tilde{}\mathbf{X} , \tag{27}$$

which is very simple indeed. For the even more special case in which all w_{ijk} are unity, we find $^+\mathbf{Z} = \tilde{}\mathbf{X}/n$, since $\tilde{}\mathbf{X}$ is already centered.

All these simplified cases also appear as special cases of the alternate solution to be presented in section 5, and - although this is not always necessary or helpful - condition (24) will from now on be adhered to, for definiteness.

4. Solutions for the individual transformations.

4.1. FULL model. When the common space $\mathbf{Z}$ is kept fixed the individual transformations can be found one after the other, because the majorizing function (4) breaks down naturally into components of the form

$$\mu^2(\mathbf{A}_k) = c_k + \text{tr } \mathbf{A}_k'\mathbf{Z}'\mathbf{V}_k\mathbf{Z}\mathbf{A}_k - 2 \text{ tr } \mathbf{A}_k'\mathbf{Z}'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \; ; \tag{28}$$

thus, setting the partial derivatives of (28) equal to zero as usual, the update ${}^*\mathbf{A}_k$ must satisfy

$$(\mathbf{Z}'\mathbf{V}_k\mathbf{Z})\ ({}^*\mathbf{A}_k) = \mathbf{Z}'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \; . \tag{29}$$

When the weight matrices are equal the coefficients of the system remain constant, so that some efficiency can be gained. Note that again the individual Guttman transforms are not needed, and that there arises no problem if the matrix on the right-hand side of (29) incidentally becomes rank deficient.

After all the ${}^*\mathbf{A}_1,\ldots,{}^*\mathbf{A}_m$ have been obtained this way they are transformed, as outlined in the previous section, so that the average cross product matrix becomes unity.

4.2. DIAGONAL model. When the $\mathbf{A}_k$ are restricted to be diagonal (the INDSCAL case), we may write $\mathbf{A}_k \equiv \mathbf{U}_k = \text{diag}(\mathbf{u}_k)$, where the p-vector $\mathbf{u}_k$, with elements u_{ak}, contains the *dimension weights*, or *saliences*, of individual k. Now a further breakdown of (28) is possible, into dimensionwise components:

$$\mu^2(\mathbf{U}_k) = c_k + \Sigma_a \, \mu^2(u_{ak}) \; , \tag{30}$$

with

$$\mu^2(u_{ak}) = (\mathbf{z}_a'\mathbf{V}_k\mathbf{z}_a)\ u^2_{ak} - 2\ (\mathbf{z}_a'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{x}_{ak})\ u_{ak} \; ; \tag{31}$$

this implies that the saliences can be obtained one after the other, within each individual, as the ratio

$${}^*u_{ak} = (\mathbf{z}_a'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{x}_{ak})/(\mathbf{z}_a'\mathbf{V}_k\mathbf{z}_a) \; , \tag{32}$$

and these may then afterwards be normalized so that their mean square across individuals becomes unity (cf. 24). Unlike other methods for fitting the diagonal model - such as the original Carroll-Chang INDSCAL procedure - there is no problem of getting "negative weights",

since the u_{ak} are estimated, not the u^2_{ak}. If ${}^*u_{ak}$ would become negative, there is simply a reflexion of the individual axis, and without changing the individual distances the absolute value can be chosen for definiteness.

4.3. REDUCED RANK model with full parametrization. The reduced rank model, or *G*eneral *E*uclidean *M*odel, assumes $\mathbf{A}_k$ to be of less than full rank, where the actual rank r is specified in advance; it is not estimated. The parameter r has a status somewhat similar to p, the dimensionality of the common space. In the present approach there is room for specifying r differently for each k, but this generality will not be pursued.

Let $\mathbf{A}_k = \mathbf{G}_k\mathbf{H}_k'$, a rank-$r$ decomposition; thus both $\mathbf{G}_k$ and $\mathbf{H}_k$ are of order p x r. For convenience of exposition, and without loosing generality, it is asssumed that $\mathbf{H}_k$ is orthonormal (i.e., $\mathbf{H}_k'\mathbf{H}_k = \mathbf{I}$; this determines $\mathbf{G}_k$ and $\mathbf{H}_k$ up to rotations in r-space). Then (28) becomes

$$\mu^2(\mathbf{G}_k,\mathbf{H}_k) = c_k + \text{tr } \mathbf{G}_k'\mathbf{Z}'\mathbf{V}_k\mathbf{Z}\mathbf{G}_k - 2 \text{ tr } \mathbf{H}_k\mathbf{G}_k'\mathbf{Z}'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k \,. \tag{33}$$

Setting the partials with respect to $\mathbf{G}_k$ equal to zero we find that the minimum is attained when ${}^*\mathbf{G}_k$ is the solution of the linear system

$$(\mathbf{Z}'\mathbf{V}_k\mathbf{Z})\,{}^*\mathbf{G}_k = \mathbf{Z}'\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k\mathbf{H}_k \,, \tag{34}$$

for any choice of $\mathbf{H}_k$. Substituting (34) in the stationary equations for minimizing the Lagrangean function that links (33) with the orthonormality constraints on $\mathbf{H}_k$, we find that ${}^*\mathbf{H}_k$ is the solution of the eigenequation

$$ {}^0\mathbf{X}_k'\mathbf{B}_k({}^0\mathbf{X}_k)\{\mathbf{Z}(\mathbf{Z}'\mathbf{V}_k\mathbf{Z})^{-1}\mathbf{Z}'\}\mathbf{B}_k({}^0\mathbf{X}_k){}^0\mathbf{X}_k\,{}^*\mathbf{H}_k = {}^*\mathbf{H}_k\Lambda, \tag{35}$$

which involves the first r eigenvectors and eigenvalues of a p x p matrix. Once ${}^*\mathbf{H}_k$ is found, ${}^*\mathbf{G}_k$ can be obtained via (34), and ${}^*\mathbf{A}_k = {}^*\mathbf{G}_k{}^*\mathbf{H}_k'$. Note that the normalization does not depend on ${}^*\mathbf{H}_k$, since ${}^*\mathbf{A}_k{}^*\mathbf{A}_k' = {}^*\mathbf{G}_k{}^*\mathbf{G}_k'$.

4.4. REDUCED RANK model with reduced individual spaces. It is possible to simplify the whole fitting process under the reduced rank assumption considerably by taking advantage of the following observation:

$$\text{if} \qquad \mathbf{X}_k = \mathbf{Z}\mathbf{G}_k\mathbf{H}_k', \qquad \text{then} \quad d_{ij}(\mathbf{X}_k) = d_{ij}(\mathbf{X}_k\mathbf{H}_k) \,. \tag{36}$$

The distances among the rows of $\mathbf{X}_k$ are equal to the distances among the rows of the *reduced individual spaces* $\mathbf{X}_k\mathbf{H}_k$. This is so because the interpoint distances depend solely on the outer

products $\mathbf{X}_k\mathbf{X}_k'$, and although it is *not* true that $\mathbf{H}_k\mathbf{H}_k' = \mathbf{I}$, it still follows from the model and the orthonormality of $\mathbf{H}_k$ that $\mathbf{X}_k\mathbf{H}_k = \mathbf{Z}\mathbf{G}_k$ and therefore $\mathbf{X}_k\mathbf{X}_k' = \mathbf{Z}\mathbf{G}_k\mathbf{G}_k'\mathbf{Z}' = \mathbf{X}_k\mathbf{H}_k\mathbf{H}_k'\mathbf{X}_k'$.

All operations can be performed in reduced dimensionality, i.e. we can work with $\mathbf{G}_k$ instead of $\mathbf{A}_k$, and with $\underline{\mathbf{X}}_k = \mathbf{X}_k\mathbf{H}_k$ instead of $\mathbf{X}_k$ itself, since:

1. STRESS remains the same for either $\mathbf{X}_k$ or $\underline{\mathbf{X}}_k$;
2. the model becomes $\underline{\mathbf{X}}_k = \mathbf{Z}\mathbf{G}_k$;
3. as remarked earlier, we have $\mathbf{A}_k\mathbf{A}_k' = \mathbf{G}_k\mathbf{G}_k'$;
4. the $\mathbf{B}_k$-matrix depends on $\mathbf{X}_k$ via the distances, so $\mathbf{B}_k(\underline{\mathbf{X}}_k) = \mathbf{B}_k(\mathbf{X}_k)$;
5. updating the common space now involves (cf. 21) $\tilde{}\mathbf{X} = 1/m \sum_k \mathbf{B}_k({}^0\underline{\mathbf{X}}_k){}^0\underline{\mathbf{X}}_k\mathbf{G}_k'$;
6. for finding ${}^*\mathbf{G}_k$, (34) reduces to $(\mathbf{Z}'\mathbf{V}_k\mathbf{Z})\,{}^*\mathbf{G}_k = \mathbf{Z}'\mathbf{B}_k({}^0\underline{\mathbf{X}}_k){}^0\underline{\mathbf{X}}_k$.

So $\mathbf{H}_k$ can be "absorbed" in the individual configurations everywhere. This is the way PROXSCAL operates, and by which quite a lot efficiency is gained: computation of the distance matrices and the preliminary updates $\mathbf{B}_k({}^0\underline{\mathbf{X}}_k){}^0\underline{\mathbf{X}}_k$ becomes shorter; for the transformation matrices the computations become identical to those of the full rank case, except for the number of columns involved (in the master program care has to be taken that r replaces p in the individual configurations and transformations, but not in the common space). Normalization is the same as before.

The question might be raised whether it is still possible to give the results in p-space after having iterated in r-space. The answer is affirmative: at any time we can use *any arbitrary* orthonormal matrix $\mathbf{H}_k$ to expand $\underline{\mathbf{X}}_k$ again, because - in contrast to the conditional case (36) - it is *always* true that $d_{ij}(\underline{\mathbf{X}}_k) = d_{ij}(\underline{\mathbf{X}}_k\mathbf{H}_k')$ if $\mathbf{H}_k$ is orthonormal, and therefore we can always put $\mathbf{X}_k = \underline{\mathbf{X}}_k\mathbf{H}_k'$, $\mathbf{A}_k = \mathbf{G}_k\mathbf{H}_k'$, and leave $\mathbf{Z}$ unchanged. This fact also shows that $\mathbf{H}_k$ can never be very informative in the first place.

4.5. DIAGONAL REDUCED RANK model. By imposing restrictions on either the DIAGONAL or the REDUCED RANK model a new, intermediate case is obtained: the DIAGONAL REDUCED RANK model. In terms of the DIAGONAL model it requires at least $p - r$ of the saliences being zero, whereas in terms of the REDUCED RANK model it requires each row and each column of $\mathbf{G}_k$ containing only one nonzero element. It is probably a useful intermediate case, because it allows us to increase p considerably without much risk for overparametrization, while still being able to maintain the INDSCAL interpretation - including its helpful unique axes property. The DIAGONAL REDUCED RANK model is an explicit formulation of the idea that the common space may be high-dimensional, whereas the individual spaces are not. As an aside it may be remarked that the DIAGONAL model can be viewed as the special case $\mathbf{A}_k = \mathbf{G}_k\mathbf{H}_k'$ with $r = p$ and $\mathbf{G}_k$ diagonal. The development of the previous section makes it clear that the rotations

$\mathbf{H}_k$ need not be explicitly estimated, and that the $\underline{\mathbf{X}}_k$ may be rotated freely without altering STRESS (but that only the identity preserves the diagonality of $\mathbf{A}_k\mathbf{H}_k$).

For the estimation of the constrained individual saliences it is convenient to start from the formulation in terms of the least squared distance function $\eta^2(\mathbf{X}_k,{}^0\mathbf{X}_k)$ as defined in (16a). Inserting the model equation and dropping reference to the previous configuration ${}^0\mathbf{X}_k$ yields

$$\eta^2(\mathbf{U}_k) = \text{tr}\ ({}^0\mathbb{X}_k - \mathbf{Z}\mathbf{U}_k)'\mathbf{V}_k({}^0\mathbb{X}_k - \mathbf{Z}\mathbf{U}_k)\ . \tag{37}$$

Suppose $\mathbb{U}_k$ is the minimizer of the full diagonal problem. It is used to split the residuals into two orthogonal components, so that (37) can be decomposed into two parts, only the second of which, and the simpler one, includes the unknown $\mathbf{U}_k$. Writing

$${}^0\mathbb{X}_k - \mathbf{Z}\mathbf{U}_k = ({}^0\mathbb{X}_k - \mathbf{Z}\mathbb{U}_k) + (\mathbf{Z}\mathbb{U}_k - \mathbf{Z}\mathbf{U}_k)\ , \tag{38}$$

and using a number of simplifications, the desired decomposition of (37) is:

$$\begin{aligned}\eta^2(\mathbf{U}_k) &= \eta^2(\mathbb{U}_k) + \text{tr}\ (\mathbb{U}_k - \mathbf{U}_k)'\mathbf{Z}'\mathbf{V}_k\mathbf{Z}(\mathbb{U}_k - \mathbf{U}_k) \\ &= \eta^2(\mathbb{U}_k) + (\mathbb{u}_k - \mathbf{u}_k)'\Psi_k(\mathbb{u}_k - \mathbf{u}_k)\ . \end{aligned} \tag{39}$$

The fact that the two components on the right-hand side of (38) are indeed orthogonal in the metric $\mathbf{V}_k$ can be verified by repeatedly using (32). Since $\mathbb{U}_k$ and $\mathbf{U}_k$ are diagonal matrices, they can be replaced in (39) by their p-vector counterparts $\mathbb{u}_k$ and $\mathbf{u}_k$ if the full matrix $\mathbf{Z}'\mathbf{V}_k\mathbf{Z}$ is also replaced by the diagonal matrix Ψ_k, containing the diagonal elements of $\mathbf{Z}'\mathbf{V}_k\mathbf{Z}$ (thus the ath diagonal element of Ψ_k is $\mathbf{z}_a'\mathbf{V}_k\mathbf{z}_a$). From (39) it is clear that the remaining problem is to select those $p - r$ elements of $\mathbb{u}_k$ for which the weighted sum of squares is minimal; these elements are made zero in the constrained solution ${}^*\mathbf{u}_k$, while the r other elements of ${}^*\mathbf{u}_k$ and $\mathbb{u}_k$ are equal. Apparently it is not only the size of an individual salience, but also the (individually weighted) dispersion $\mathbf{z}_a'\mathbf{V}_k\mathbf{z}_a$ of a common axis that contributes to the decision whether or not axis a is "in" for individual k.

5. Dimensionwise solution for the common space coordinates.

In section 3 it was argued that the *simultaneous* solution for all np coordinates of the common space is not efficient in most practical cases; in addition, it would be useful to have an approach that easily generalizes to the situation where subsets of the coordinates are restricted. This section will outline such an approach for a dimensionwise split of the parameters. Thus the

subproblem of minimizing the majorizing function over $\mathbf{Z}$ for fixed $\mathbf{A}_1,...,\mathbf{A}_m$ is again divided into p subproblems of smaller size, through which we have to cycle at least once, perhaps a number of times, but not necessarily until convergence. For pointwise or still other partitions of $\mathbf{Z}$ completely analogous procedures can be developed.

Consider the ath column of $\mathbf{Z}$, denoted as $\mathbf{z}_a$, and the auxiliary matrix $\mathbf{P}_a$, defined as

$$\mathbf{P}_a \equiv \mathbf{Z} - \mathbf{z}_a\mathbf{e}_a' , \quad \text{so that } \mathbf{Z} = \mathbf{P}_a + \mathbf{z}_a\mathbf{e}_a' ; \tag{40}$$

here $\mathbf{e}_a$ is the ath column of the identity matrix. $\mathbf{P}_a$ is $\mathbf{Z}$, with ath column replaced by zeros. Now the majorizing function (20), expressed as a function of $\mathbf{z}_a$ only, becomes

$$\begin{aligned}\mu^2(\mathbf{z}_a) = c_* &+ \text{tr } (\mathbf{P}_a + \mathbf{z}_a\mathbf{e}_a')'\{1/m \textstyle\sum_k \mathbf{V}_k(\mathbf{P}_a + \mathbf{z}_a\mathbf{e}_a')\mathbf{C}_k\} - \\ &- 2 \text{ tr } (\mathbf{P}_a + \mathbf{z}_a\mathbf{e}_a')'(\tilde{}\mathbf{X}) .\end{aligned} \tag{41}$$

In order to simplify (41), let us define the reweighted mean weight matrix $\mathbf{V}_a$ and the corrected vector of constant terms ${}^*\mathbf{x}_a$ as

$$\mathbf{V}_a \equiv 1/m \textstyle\sum_k (\mathbf{e}_a'\mathbf{C}_k \mathbf{e}_a)\mathbf{V}_k , \tag{42a}$$

$${}^*\mathbf{x}_a \equiv \tilde{}\mathbf{X}\mathbf{e}_a - 1/m \textstyle\sum_k \mathbf{V}_k\mathbf{P}_a\mathbf{C}_k\mathbf{e}_a . \tag{42b}$$

Using these definitions, absorbing all constant scalars into a single one, $*c_*$, and with the target vector $\mathbb{x}_a$ defined as any solution of the linear system ${}^*\mathbf{x}_a = \mathbf{V}_a \mathbb{x}_a$, (41) becomes equivalent to

$$\mu^2(\mathbf{z}_a) = *c_* + (\mathbb{x}_a - \mathbf{z}_a)'\mathbf{V}_a(\mathbb{x}_a - \mathbf{z}_a) . \tag{43}$$

i.e., a projection problem in the metric $\mathbf{V}_a$. If $\mathbf{z}_a$ is unrestricted, the conditionally optimal ${}^+\mathbf{z}_a$ simply is $\mathbb{x}_a$ (there is no need for projection in this case, since $\mathbf{z}_a$ can be everywhere in the space of centered n-vectors). If $\mathbf{z}_a$ is restricted to be in some subspace, the solution of the projection problem is standard. However, if $\mathbf{z}_a$ is restricted to be in some more complicated subset of the R^n, e.g. a polyhedral convex cone, then the non-diagonality of $\mathbf{V}_a$ makes the projection more complicated as well. A solution based on majorization will be discussed in a forthcoming PROXSCAL note.

In (42b) the ath column of $\tilde{}\mathbf{X}$ is corrected with contributions from all other columns of $\mathbf{Z}$ (indeed not $\mathbf{z}_a$ itself, since the vector $\mathbf{C}_k\mathbf{e}_a$ combines the columns of $\mathbf{P}_a$, the ath of which contains zeros). If the model is DIAGONAL, then $\mathbf{C}_k$ is diagonal, so that $\mathbf{C}_k\mathbf{e}_a$ contains a nonzero element on the ath position only, and therefore $\mathbf{P}_a\mathbf{C}_k\mathbf{e}_a$ and consequently the whole correction term vanishes. Since $\mathbf{V}_a$ then also reduces to one of the block-diagonal parts of (23),

the present procedure becomes equivalent to the simultaneous solution of section 3 for the DIAGONAL model. One could say that the complexity of computation under the other models is reduced to the order of complexity for the DIAGONAL model. One could also say that in the unrestricted case the present scheme is essentially an application of the Gauss-Seidel method for solving linear equations. In the restricted case it is an application of a generalization of the Gauss-Seidel method, in psychometrics usually called Alternating Least Squares.

6. Discussion.

The approach to individual differences scaling developed in this report has introduced a number of improvements and innovations compared to the existing literature. First and foremost, it comprises the first detailed treatment of convergent algorithmic procedures in the STRESS framework, at a level which permits others too - it is hoped - to implement them with some efficiency in a high level language. All models except for the DIAGONAL REDUCED RANK model are available in PROXSCAL. The possibility to use data weights makes it very easy to adapt the procedures for missing data, nonmetric transformations, and unfolding options, all of which are available in PROXSCAL. The material in sections 4.4, 4.5, and 5 is new, and suggests how to handle further extensions, e.g. the PROXSCAL option EXTERNAL, which imposes restrictions on $\mathbf{Z}$ of the form $\mathbf{Z} = \mathbf{ER}$, with $\mathbf{E}$ some set of *external variables*, of prespecified measurement level, and $\mathbf{R}$ a set of *regression weights*. The EXTERNAL option is a STRESS implementation and generalization of CANDELINC (Carroll *et al.* , 1976), using ideas of Meulman and Heiser (1984).

In their presentation of the SMACOF "algorithm model" De Leeuw and Heiser (1980) discern two basic steps: finding the Guttman transform(s) first, and next projecting it (them) onto the space of admissible configurations. De Leeuw (1980) also uses such a two-step scheme. The present report has primarily used the equivalent concept of repeatedly minimizing the majorizing function. This is only a matter of emphasis. When the configuration(s) are restricted, the Guttman transform is always - both conceptually and practically - a move in the wrong direction, and its computation should therefore preferably be avoided. It does give the correct location of the unrestricted minimum of the majorizing function, and thus it is a useful theoretical anchorpoint to define all configurations that would decrease the STRESS in a region adjacent to ${}^0\mathbf{X}_1,\ldots,{}^0\mathbf{X}_m$. (Because all arguments remain valid if attention is restricted to a single constrained configuration, reference to k is dropped from here on). From (10) we see that any ${}^*\mathbf{X}$ for wich $\mu^2({}^*\mathbf{X},{}^0\mathbf{X}) < \mu^2({}^0\mathbf{X},{}^0\mathbf{X})$ falls into that region, and using (15) and (16a) this implies that ${}^*\mathbf{X}$ must satisfy

$$\text{tr}\ ({}^0\mathbb{X} - {}^*\mathbf{X})'\mathbf{V}({}^0\mathbb{X} - {}^*\mathbf{X}) < \text{tr}\ ({}^0\mathbb{X} - {}^0\mathbf{X})'\mathbf{V}({}^0\mathbb{X} - {}^0\mathbf{X})\ . \tag{44}$$

Thus all $^*\mathbf{X}$ within a hyperellipsoid around $^0\mathbb{X}$ and through $^0\mathbf{X}$ are satisfactory updates. Although at any one step it is not necessary that $^*\mathbf{X}$ is *also* in the feasible region, it must be there at the point of convergence, and the safest way to achieve this is to keep the whole succession of updates within the feasible region. A number of these concepts are illustrated in

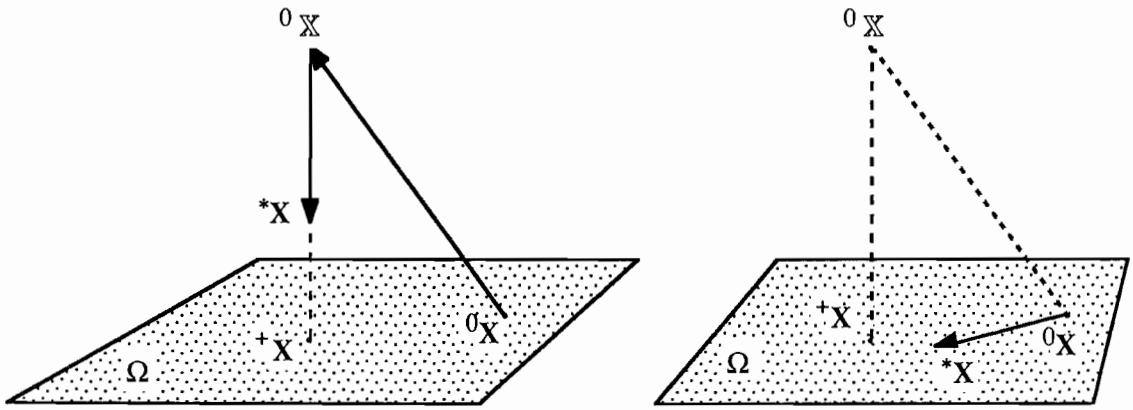

Figure 1.

Figure 1. All configurations of interest are pictured in three-dimensional space, and the feasible region Ω as part of a plane. For convenience let us assume that $\mathbf{V}$ is the identity (up to centering), so that the hyperellipsoid becomes a sphere. In Figure 1(a), we first move from $^0\mathbf{X}$ to $^0\mathbb{X}$, and $^+\mathbf{X}$ is the projection of $^0\mathbb{X}$ onto Ω; going from $^0\mathbf{X}$ to $^+\mathbf{X}$ is the complete SMACOF step. Suppose we would choose $^*\mathbf{X}$ as an update (e.g. by changing $^0\mathbb{X}$ such that it satisfies only a subset of the constraints), then STRESS would decrease, since the sphere around $^0\mathbb{X}$ going through $^0\mathbf{X}$ includes $^*\mathbf{X}$. However, whether or not we are still OK the next iteration depends on $^*\mathbb{X}$, the Guttman transform of $^*\mathbf{X}$; if the sphere around $^*\mathbb{X}$ through $^*\mathbf{X}$ would not intersect with Ω, then there would not exist a STRESS decreasing *and* feasible update anymore. Because the convergence of the whole process is monitored by the rate of change of STRESS, not staying within the feasible region impies running the risk of stopping too early.

In Figure 1(b) the path is sketched along which we normally proceed: directly from $^0\mathbf{X}$ to $^*\mathbf{X}$ by partly minimizing the majorizing function, respecting all constraints. Note that $^*\mathbf{X} - {}^0\mathbf{X}$ will generally not point exactly towards $^+\mathbf{X}$, since fixing subsets of the parameters while improving others implies a "zigzagging" path towards the minimum. The projection of a sphere with center $^0\mathbb{X}$ onto a plane is a circle with center $^+\mathbf{X}$ in Ω, which also goes through $^0\mathbf{X}$. As long as $^*\mathbf{X}$ is kept within that circle regular convergence is guaranteed. This, in turn, is achieved when we make sure that the values of the majorizing function never increase.

7. References.

Bloxom, B. (1978). Constrained multidimensional scaling in N spaces. *Psychometrika, 43*, 397-408.

Carroll, J.D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283-319.

Carroll, J.D., Green, P.E. and Carmone, F.J. (1976). CANDELINC (CANonical DEcompositions with LINear Constraints). *Paper presented at the 84th annual convention of the American Psychological Association*, San Francisco.

Carroll, J.D. and Wish, M. (1974). Models and methods for three-way multidimensional scaling. In: D.H. Krantz *et al.* (Eds.), *Contemporary developments in mathematical psychology, Vol 2: Measurement, psychophysics, and neural information processing.* San Francisco: Breman, 57-105.

De Leeuw, J. (1980). Majorization algorithms for individual differences in multidimensional scaling. *Paper presented at the symposium "Multidimensional scaling and interindividual differences" at the 22nd International Congress of Psychology*, Leipzig, GDR.

De Leeuw, J. and Heiser, W.J. (1977). Convergence of correction matrix algorithms for multidimensional scaling. In: J. Lingoes (Ed.), *Geometric representations of relational data.* Ann Arbor: Mathesis Press, 735-752.

De Leeuw, J. and Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In: P.R. Krishnaiah (Ed.), *Multivariate Analysis, Vol V.* Amsterdam: North-Holland, 501-522.

Horan, (1969). Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika, 34*, 139-165.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1-28.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika, 29*, 115-129.

Meulman, J. and Heiser, W.J. (1984). Constrained multidimensional scaling: more directions than dimensions. In: T. Havránek et al. (Eds.), *COMPSTAT 1984, Proceedings in Computational Statistics.* Vienna: Physica Verlag, 137-142.

Stoop, I., and De Leeuw, J. (1982). How to use SMACOF-1B. *Research Report*, Leiden: Department of Data Theory.

Stoop, I., Heiser, W.J., and De Leeuw, J. (1981). How to use SMACOF-1A. *Research Report* , Leiden: Department of Data Theory.

Takane, Y., Young, F.W. and De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42*, 7-67.

Young, F.W. (1984). The general Euclidean model. In: H.G. Law *et al.* (Eds.), *Research methods for multimode data analysis.* New York: Praeger, 440-469.