On the Procrustean analogue of individual differences scaling (INDSCAL)

Steffen Unkel*[1], John C. Gower[2] and Nickolay T. Trendafilov[2]

[1]Department of Statistics
Ludwig-Maximilians-University Munich, Germany

[2]Department of Mathematics and Statistics
The Open University, Milton Keynes, United Kingdom

May 3, 2013

Abstract

In this paper, individual differences scaling (INDSCAL) is revisited, considering INDSCAL as being embedded within a hierarchy of individual difference scaling models. We explore the members of this family, distinguishing (i) models, (ii) the role of identification and substantive constraints, (iii) criteria for fitting models and (iv) algorithms to optimise the criteria. Model formulations may be based either on data that are in the form of proximities or on configurational matrices. In its configurational version, individual difference scaling may be formulated as a form of generalized Procrustes analysis. Algorithms are introduced for fitting the new models. An application from sensory evaluation illustrates the performance of the methods and their solutions.

Key words: INDSCAL, Multidimensional scaling, Configuration matrices, Procrustes problems, Proximities, Three-way data, SMACOF.

* Correspondence should be addressed to: Steffen Unkel, Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstrasse 33, 80539 Munich, Germany (e-mail:Steffen.Unkel@informatik.med.uni-giessen.de).

1 Setting the scene

As its name implies, Individual Differences Scaling is concerned with analysing and comparing the difference between multidimensional scaling (MDS) solutions pertaining to the same n stimuli observed for K individuals. For further information about MDS, see Borg and Groenen (2005) and Cox and Cox (2001). There are many ways in which individual MDS may be compared but all summarise the individual scalings by some kind of group average, which may be represented and visualised by a set of n points, in a similar way as for individual MDS. Different approaches depend on how the group average is related to the individual scalings and we shall discuss some of these in the following. First, to fix ideas, we shall rehearse the original approach of Carroll and Chang (1970).

1.1 Individual differences scaling – Carroll and Chang (1970)

Originally developed in the psychometric literature to explain the relationship between subjects' differential cognition of a set of stimuli, *IN*dividual *D*ifferences *SCAL*ing (INDSCAL) has found applications in various disciplines, ranging from the social sciences to chemometrics. In this paper, we use INDSCAL both to denote the method of Carroll and Chang (1970) and to distinguish it from other methods of individual differences scaling. We consider K given $n \times n$ matrices of distances or dissimilarities $\mathbf{D}_k$ $(k = 1, \ldots, K)$. INDSCAL is a method that aims to represent the matrices $\mathbf{D}_1, \ldots, \mathbf{D}_K$, one for each individual, in terms of a group average matrix $\mathbf{G}$, defining coordinates of the n stimuli on each of R pre-specified dimensions, together with diagonal weight matrices $\mathbf{W}_k$ $(k = 1, \ldots, K)$. The weights $\mathbf{W}_k$ (often called saliences) refer to each of the R dimensions and vary over the K individual sets. Carroll and Chang (1970) first put the distances or dissimilarities into inner-product form by the double centring operation (Torgerson, 1958, pp. 254-259):

$$\mathbf{B}_k = -\frac{1}{2}\mathbf{J}_n(\mathbf{D}_k \odot \mathbf{D}_k)\mathbf{J}_n \ , \tag{1}$$

where $\odot$ denotes the elementwise (Hadamard) matrix product and $\mathbf{J}_n = \mathbf{I}_n - n^{-1}\mathbf{E}_n$ is the $n \times n$ centring matrix with $\mathbf{E}_n$ being an $n \times n$ matrix of ones. The INDSCAL model approximates each centred matrix $\mathbf{B}_k$ by

$$\mathbf{B}_k \approx \mathbf{G}\mathbf{W}_k^2\mathbf{G}^\top \ .$$

Specifically, the INDSCAL problem seeks for $(\mathbf{G}, \mathbf{W}_1^2, \ldots, \mathbf{W}_K^2)$ such that the model fits the data in a least squares sense. That is, INDSCAL minimises

$$\sum_{k=1}^{K} ||\mathbf{B}_k - \mathbf{G}\mathbf{W}_k^2\mathbf{G}^\top||^2 \ , \tag{2}$$

where $||\mathbf{A}|| = \sqrt{\text{trace}(\mathbf{A}^\top\mathbf{A})}$, $\mathbf{G} \in \mathbb{R}^{n\times R}$ has full column rank R and $\mathbf{W}_k^2$ are diagonal and non-negative $R \times R$ matrices. The restriction to non-singular weights ensures that the INDSCAL model has the property that solutions cannot be rotated without changing the solution in an essential way. However, note that $\mathbf{G}\mathbf{W}_k = (\mathbf{G}\mathbf{L})(\mathbf{L}^{-1}\mathbf{W}_k)$ for any appropriate diagonal matrix $\mathbf{L}$. That is, if $\mathbf{G}$ is stretched by $\mathbf{L}$ and the weights in $\mathbf{W}_k$ are subjected to the inverse transformation, the product remains the same. To handle this indeterminacy, one can assign one of the following *identification* constraints without substantively changing the problem:

$$\sum_{k=1}^{K} \mathbf{W}_k^2 = K\mathbf{I}_R \tag{3}$$

or

$$\text{diag}(\mathbf{G}^\top\mathbf{G}) = \mathbf{I}_R \ , \tag{4}$$

i.e. requiring $\mathbf{G}$ to be an oblique matrix. Thus, if $\mathbf{G}\mathbf{W}_k$ does not already satisfy (3) or (4), we can always choose $\mathbf{L}$ accordingly.

Minimizing (2) has no direct analytical solution. The standard numerical solution is given by an alternating least squares (ALS) algorithm, called CANDECOMP (Carroll and Chang, 1970). The two appearances of $\mathbf{G}$ in (2) may be represented by different matrices, say $\mathbf{G}$ and $\mathbf{H}$. Then, (2) is called the CANDECOMP function. Optimization is carried out on $\mathbf{G}$ and $\mathbf{H}$ independently, along with $\mathbf{W}_1^2, \ldots, \mathbf{W}_K^2$. The belief is that after convergence of CANDECOMP we have $\mathbf{G} = \mathbf{H}$ as is required by the model. This is known as the "symmetry requirement" and usually seems to be satisfied in practice, although there is no general proof of the claim. On the contrary, it has been shown by ten Berge and Kiers (1991) that the CANDECOMP algorithm can produce asymmetric INDSCAL solutions ($\mathbf{G} \neq \mathbf{H}$) even for positive semi-definite data $\mathbf{B}_k$. Also, one must hope that the solutions obtained for $\mathbf{W}_1^2, \ldots, \mathbf{W}_K^2$ have non-negative diagonal elements throughout.

This is not always the case in practice. ten Berge et al. (1993) avoided negative saliences by imposing non-negativity constraints and preserved the symmetry of the solution by developing an algorithm named SYMPRES.

In the loss function (2), $\mathbf{G}$ is assumed to be of full column rank. The parameter set of all $n \times R$ matrices $\mathbf{G}$ with full column rank R is a *non-compact* set (Absil et al., 2008). Optimisation over a non-compact set may cause computational problems. For example, Stegeman (2007) has pointed out that in certain cases the least squares criterion (2) to be minimized in INDSCAL does not have a minimum, but only an infimum, within the parameter space. In such situations, CANDECOMP is bound to produce a sequence of updates of parameters which is degenerate in the sense that some columns of $\mathbf{G}$ are highly correlated and some of the diagonal elements of $\mathbf{W}_k^2$ are arbitrarily large. These degenerate solutions are hard to interpret. As a remedy, one is tempted to minimise (2) subject to

$$\mathbf{G}^\top \mathbf{G} = \mathbf{I}_R \ . \tag{5}$$

The column-wise orthonormal matrix $\mathbf{G}$ forms a *compact* set and then the INDSCAL minimisation problem (2) has an attainable solution. Indeed, Krijnen et al. (2008) have shown that if one considers an INDSCAL model which constrains the group stimulus space to be column-wise orthonormal, then there is no degeneracy problem. Imposing $\mathbf{G}^\top \mathbf{G} = \mathbf{I}_R$ is also computationally more efficient (Takane et al., 2010). Kroonenberg (1983) calls the method that minimises (2) subject to (5) "orthonormal INDSCAL". The orthonormal INDSCAL model has been considered by several authors (ten Berge et al., 1988; Kiers, 1989; Trendafilov, 2004, 2012; Takane et al., 2010).

However, $\mathbf{G}^\top \mathbf{G} = \mathbf{I}_R$ is a *substantive* constraint and not just an identification constraint as are (3) or (4). Despite its superficial similarity to (4), (5) is a strong constraint that substantially alters the model. In fact, minimising (2) subject to $\mathbf{G}^\top \mathbf{G} = \mathbf{I}_R$ turns INDSCAL into a least squares version of common components analysis (Flury, 1988). We have reservations on the wisdom of applying substantive constraints on model parameters just for the purpose of algorithmic convenience. Moreover, if a degenerate solution does arise, it suggests to us that one should not be fitting an INDSCAL type model in the first place - the data are not compatible with the model. In favour of the use of (5), Takane et al.

(2010) give empirical evidence that the original INDSCAL and orthonormal INDSCAL for some applications give (i) very similar results and (ii) the original INDSCAL model yields nearly orthogonal solutions on $\mathbf{G}$ without explicitly requiring the orthogonality. But it seems that this observation is quite a good argument for not imposing the constraint - i.e. usually (5) is not a constraint of any consequence. It seems that many, perhaps all, of these difficulties are related to Heywood cases; that is, when parameter estimates of some saliences occur on the boundary of the feasible region, so generating infimum solutions. Other difficulties may be properties of using a least squares criterion. Yet others may be failings in the INDSCAL model itself when fitting to degenerate data; after all even linear least squares will fail with reduced-rank data. We have already seen that CANDECOMP may arrive at asymmetric solutions; an algorithm that does not have inbuilt asymmetry could only deliver symmetric results and may do better in other respects. Our feeling is that while it is interesting to investigate aberrant behaviour, rather than searching for fixes, it is often more rewarding to seek why an aberration has occurred in the first place. In summary:

(i) Distinguish between the model, the fitting criterion and the algorithm for optimising the fit.

(ii) Identification constraints are benign but substantive constraints need justification, because they are a fundamental part of the model.

(iii) If algorithms fail with some data, the problem may lie with the data - informative diagnostics may be more useful than fixes designed to deal with aberrant behaviour.

1.2 Other Individual Differences Scaling models

In the above discussion of INDSCAL we have introduced the dissimilarity matrices $\mathbf{D}_k$ forming the data, the derived inner-product matrices $\mathbf{B}_k$ and the group average $\mathbf{G}$. The reasons for using the inner-products $\mathbf{B}_k$ in INDSCAL rather than the more fundamental data $\mathbf{D}_k$ are probably because (i) $\mathbf{B}_k$ has better algebraic properties than does $\mathbf{D}_k$ and (ii) $\mathbf{B}_k$ provides a firm link with classical scaling. In Section 2 of this paper we discuss direct analysis of $\mathbf{D}_k$.

Even more directly, we may regard the data as configuration matrices $\mathbf{X}_k \in \mathbb{R}^{n\times R}$ $(k = 1, \ldots, K)$ of n points with coordinates in R dimensions. The configurations $\mathbf{X}_k$ may be given directly, as for example by measuring landmark coordinates in shape studies. Alternatively, they may be derived from MDS. Or a set of data matrices $\mathbf{Y}_k \in \mathbb{R}^{n\times p}$ $(k = 1, \ldots, K)$ for n cases on p variables may be transformed to similarity matrices and then converted into configurations $\mathbf{X}_k$ by using any MDS method (possibly nonmetric). Whatever the derivation of the configurations, they may be compared directly with a group average $\mathbf{G}$, thus viewing individual differences as a form of generalized Procrustes analysis (see e.g. Gower and Dijksterhuis, 2004). The Procrustean analogue of INDSCAL has robust potential as it is formulated in terms of first-order matrices $\mathbf{X}_k$ rather than second-order matrices $\mathbf{B}_k$. In the Procrustean version of individual scaling, rotational indeterminacy in the configurations can be accommodated by considering $\mathbf{X}_k\mathbf{Q}_k$, where $\mathbf{Q}_k$ are orthogonal matrices. We introduce algorithms to provide estimates for the model unknowns $\mathbf{G}$, $\mathbf{Q}_k$ and $\mathbf{W}_k$.

As we have seen, the CANDECOMP algorithm plays a prominent part in Carroll and Chang's (1970) development of INDSCAL. Harshman (1978, 1982) developed an equivalent algorithm, DEDICOM, explicitly for fitting models either $\mathbf{ARB}^\top$ or $\mathbf{ARA}^\top$, where $\mathbf{A}$ and $\mathbf{B}$ are $n\times R$ matrices and $\mathbf{R}$, of order $R\times R$, is usually a non-diagonal asymmetric matrix. The background is in factor analysis. The symmetric form $\mathbf{ARA}^\top$ includes the INDSCAL model $\mathbf{GW}^2\mathbf{G}^\top$. The algebraic properties and algorithms for fitting a three-mode decomposition of the form $z_{ijk} = \sum_{p,q,r}^{P,Q,R} a_{ip}b_{jq}c_{kr}$ $(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K)$ have generated a substantial literature (Kroonenberg, 2008). This decomposition subsumes a variety of models, including INDSCAL, that might be fitted to data. However, as we have seen, INDSCAL fits rather awkwardly into this framework, because (i) we are dealing with symmetric matrices and (ii) of the problem of guaranteeing symmetric solutions. With configurational matrices, $\mathbf{X}_k$, a greater problem is that there is no reason why equivalent columns of e.g. $\mathbf{X}_1$ refer to the same dimensions as those in $\mathbf{X}_2$, as would be admissible with data matrices. With all these difficulties, we prefer to work directly with K symmetric matrices or K configuration matrices, which ever is the case.

The remainder of the paper is organized as follows. In Section 2, the INDSCAL model is

viewed as being embedded within a hierarchy of models, each layer with possible formulations based on proximity and configuration matrices. Algorithms for fitting the orthogonal Procrustean analogue of INDSCAL are considered in Section 3. A configuration version of INDSCAL with projection matrices is considered in Section 4. In Section 5, the methods developed in this paper are applied to food science data. Concluding comments are given in Section 6.

2 A hierarchy of models

We have discussed above the main features of individual differences scaling models, where K sets of data are presented either as K distance matrices $\mathbf{D}_k$, K inner-product matrices $\mathbf{B}_k$ or K sets of configuration matrices $\mathbf{X}_k$ $(k = 1, \ldots, K)$. The data for the K individuals are modelled in terms of a common group average $\mathbf{G}$, possibly modified by dimension weights $\mathbf{W}_k$ (the saliences) and possibly $\mathbf{G}$ is constrained to be an orthogonal matrix. The rows of $\mathbf{G}$ give the coordinates that generate the matrix $\mathbf{\Delta}$ of squared distances, which we shall denote by $\mathbf{G} \rightarrow \mathbf{\Delta}$. The least squares criteria that stem from the combination of different forms of data and different models are summarised in Table 1. Several provisos

Table 1: The relationship between (i) individual differences models (ii) types of data and (iii) least squares criteria.

	Criteria		
Model	Inner-product	Distance	Procrustean
1. $\mathbf{G}$	$\sum_{k=1}^{K} \|\|\mathbf{B}_k - \mathbf{G}\mathbf{G}^\top\|\|^2$	$\sum_{k=1}^{K} \|\|\mathbf{D}_k - \mathbf{\Delta}\|\|^2$	$\sum_{k=1}^{K} \|\|\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}\|\|^2$
2. $\mathbf{G}\mathbf{W}_k$	$\sum_{k=1}^{K} \|\|\mathbf{B}_k - \mathbf{G}\mathbf{W}_k^2\mathbf{G}^\top\|\|^2$	$\sum_{k=1}^{K} \|\|\mathbf{D}_k - \mathbf{\Delta}_k\|\|^2$	$\sum_{k=1}^{K} \|\|\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}\mathbf{W}_k\|\|^2$
3. $\mathbf{G}\mathbf{W}_k$, $\mathbf{G}^\top\mathbf{G} = \mathbf{I}$	$\sum_{k=1}^{K} \|\|\mathbf{B}_k - \mathbf{G}\mathbf{W}_k^2\mathbf{G}^\top\|\|^2$	$\sum_{k=1}^{K} \|\|\mathbf{D}_k - \mathbf{\Delta}_k\|\|^2$	$\sum_{k=1}^{K} \|\|\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}\mathbf{W}_k\|\|^2$

and variants that pertain to this table have to be considered and are briefly reviewed immediately below and, when appropriate, are referred to in the subsequent text.

Proviso 1. Of course, we need not use least squares and, when we do, we may prefer to use weighted least squares. Either the rows and/or columns of the matrices may be weighted. Such weights are unrelated to the salient weights $\mathbf{W}_k$, which require estimation.

Proviso 2. Usually the weights $\mathbf{W}_k$ are required to be diagonal with positive values.

Sometimes they may be positive definite symmetric matrices, when they are termed idiosyncratic.

Proviso 3. The fitted matrices $\mathbf{\Delta}$ and $\mathbf{\Delta}_k$ may contain distances or squared distances.

Proviso 4. Many algorithms to fit the criteria exist. Apart from the simple model $\mathbf{G}$ using the inner-product criterion, all need some kind of iterative procedure.

Proviso 5. Often the data will have been subjected to some form of prescaling. For example, observed data $\mathbf{Y}_k$ may be transformed to $\mathbf{X}_k$ or derived from a multidimensional scaling. This can be a convenient way of reducing dimensionality before individual differences are assessed.

Proviso 6. The dimensionality R of the group average $\mathbf{G}$ is often specified as part of the model. In other cases in Layer 1 of Table 1 a full p dimensional group average matrix is found and approximated in R dimensions, usually by using Principal Components. Another possibility discussed below is to replace the orthogonal matrices $\mathbf{Q}_k$ by column-wise orthonormal projection matrices $\mathbf{P}_k$ with R columns.

Proviso 7. The orthogonal matrices $\mathbf{Q}_k$ that occur in the Procrustean criteria, are intended to allow for the arbitrary orientations that are automatically subsumed by the distance and inner-product criteria. This allowance is essential to account for the arbitrary relative orientations of sets of individual scaling solutions $\mathbf{X}_k$, which we refer to as configurations, which do not match across columns. When the columns of $\mathbf{X}_k$ refer to repeated observed variables, rather than configurations, then the orthogonal matrices $\mathbf{Q}_k$ may have a factor structure interpretation. Alternatively, $\mathbf{Q}_k$ may be deleted altogether from the Procrustean criteria.

The models discussed in Table 1 form a sequence of layers each with formulations based on inner-products, proximity and configuration matrices. The layers are ordered so that each is a constrained version of its predecessor. These models are discussed in turn.

First layer (Model $\mathbf{G}$)

In this model, every individual is referred to the same group average, so strictly speaking this layer does not model individual differences. The inner-product version of this most

simple model (see Table 1) is to solve

$$\min \sum_{k=1}^{K} ||\mathbf{B}_k - \mathbf{G}\mathbf{G}^\top||^2 \ . \tag{6}$$

Since $\sum_{k=1}^{K} ||\mathbf{B}_k - \mathbf{G}\mathbf{G}^\top||^2 = \sum_{k=1}^{K} ||\mathbf{B}_k - \bar{\mathbf{B}}||^2 + K||\overline{\mathbf{B}} - \mathbf{G}\mathbf{G}^\top||^2$, the solution to (6) is determined solely by minimising $||\overline{\mathbf{B}} - \mathbf{G}\mathbf{G}^\top||^2$, which is classical scaling of $\overline{\mathbf{B}}$, with the usual R-dimensional approximation.

In the distance matrix version of the most simple model $\mathbf{G}$, we have to solve

$$\min \sum_{k=1}^{K} ||\mathbf{D}_k - \boldsymbol{\Delta}||^2 \ . \tag{7}$$

Since $\sum_{k=1}^{K} ||\mathbf{D}_k - \boldsymbol{\Delta}||^2 = \sum_{k=1}^{K} ||\mathbf{D}_k - \overline{\mathbf{D}}||^2 + K||\overline{\mathbf{D}} - \boldsymbol{\Delta}||^2$, the solution to (7) is determined solely from the least squares scaling of the average dissimilarity matrix over the K sets, minimizing STRESS: $\sum_{k=1}^{K} ||\overline{\mathbf{D}} - \boldsymbol{\Delta}||^2$ over $\mathbf{G} \rightarrow \boldsymbol{\Delta}$, where $\mathbf{G}$ is specified to have R dimensions. Recall that least squares scaling satisfies an analysis of variance $||\overline{\mathbf{D}}||^2 = ||\boldsymbol{\Delta}||^2 + ||\overline{\mathbf{D}} - \boldsymbol{\Delta}||^2$, so that $||\boldsymbol{\Delta}||^2$ is maximal. A majorization algorithm is available in the SMACOF program (De Leeuw, 1977; De Leeuw and Mair, 2009). The corresponding criterion where distances/dissimilarities are treated as squared distances/dissimilarities is known as SSTRESS, which is minimized by the ALSCAL program (Takane et al., 1977), now satisfying an analysis of variance on squared distances. In both cases, the corresponding $\mathbf{G}$ is known up to an arbitrary orthogonal transformation. For identification purposes, $\mathbf{G}$ can be represented relative to its principal axes, though normally this will be superfluous because SMACOF and ALSCAL will have provided R-dimensional solutions suitable for visualisations without any further modification.

In the Procrustean version of the most simple model $\mathbf{G}$, we have generalized orthogonal Procrustes analysis (see e.g. Gower and Dijksterhuis, 2004), which iteratively solves:

$$\min \sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}||^2 \ , \tag{8}$$

where $\mathbf{Q}_k$ are orthogonal $R \times R$ matrices ($k = 1, \ldots, K$). In (8), $\mathbf{G} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}_k\mathbf{Q}_k$ is the average configuration of the initial matrices $\mathbf{X}_1, \ldots, \mathbf{X}_K$ after transformation by $\mathbf{Q}_1, \ldots, \mathbf{Q}_K$. The optimization problem (8) is equivalent to minimizing $||\mathbf{X}_k\mathbf{Q}_k\mathbf{Q} - \mathbf{G}\mathbf{Q}||^2$

for arbitrary orthogonal $R \times R$ matrices $\mathbf{Q}$, so raises an identification issue. Again, this is usually settled by choosing $\mathbf{Q}$ so that $\mathbf{GQ}$ is referred to its principal axes (see e.g. Gower and Dijksterhuis, 2004) This approach allows the simultaneous visualisation in r dimensions of the individual configurations, together with their group average.

Second layer (Model $\mathbf{GW}_k$)

The inner-product version of this weighted Euclidean model (Borg and Groenen, 2005) coincides with that of Carroll and Chang (1970), which we have already discussed in Subsection 1.1 of this paper. Carroll and Chang (1972) generalized the weighted Euclidean model to a *generalized Euclidean* model, which they named IDIOSCAL (*I*ndividual *DI*ffererences in *O*rientation *SCAL*ing) model. In IDIOSCAL the matrix $\mathbf{G}$ is defined as in INDSCAL, but the $\mathbf{W}_k$ are symmetric positive definite or semi-definite matrices. Whereas INDSCAL provides unique axes, the IDIOSCAL model allows for rotations of the group stimulus space. When $\mathbf{W}_k$ is restricted to be diagonal, IDIOSCAL reduces to INDSCAL.

The distance version of the model $\mathbf{GW}_k$ seeks a solution to the following optimization problem:

$$\min \sum_{k=1}^{K} ||\mathbf{D}_k - \boldsymbol{\Delta}_k||^2 \ . \tag{9}$$

The routine SMACOF for individual differences solves (9) by using a majorization approach (De Leeuw and Heiser, 1980; De Leeuw and Mair, 2009) to find a group space $\mathbf{G}$ and dimension weights $\mathbf{W}_k$ associated with K dissimilarity matrices $\mathbf{D}_k$ $(k = 1, \ldots, K)$. The weighted Euclidean model can be implemented in several ways. An elaborated overview of algorithms is given in Chapter 22 of Borg and Groenen (2005).

The configurational version of $\mathbf{GW}_k$ solves the following Procrustes problem (Gower and Dijksterhuis, 2004, p. 171):

$$\min \sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{Q}_k - \mathbf{GW}_k\|^2 \ , \tag{10}$$

expressing that each (rotated) configuration approximates the weighted group average, subject to the constraint (3) or (4). In this Procrustean formulation, the matrices $\mathbf{X}_k$ may be given or will have been derived from some form of MDS. In that case, one possibility is to use classical scaling of double centred inner-product matrices $\mathbf{B}_k$ but this

would detract from the robust potential of (10) in working in terms of first order matrices $\mathbf{X}_k$ rather than second order matrices $\mathbf{B}_k$. Therefore, to reinforce robustness it would be better to use a more robust form of metric MDS such as least squares scaling, deriving the configuration $\mathbf{X}_k$ from the K distance or dissimilarity matrices $\mathbf{D}_k$, by minimizing STRESS.

In the term $\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}\mathbf{W}_k$, the orthogonal matrix $\mathbf{Q}_k$ preserves the distances approximated by whatever MDS method has been used to generate the configuration matrices $\mathbf{X}_k$, so conforming with the notion of fitting configurational distance matrices. The dimension scaling (i.e. saliences) is attached to the group average matrix and this conforms with the INDSCAL usage. This formulation is equivalent to a model term $\mathbf{X}_k - \mathbf{G}\mathbf{W}_k\mathbf{Q}_k^\top$ where the scaled group average is rotated to fit the configurations. This differs from the simple variant $\mathbf{X}_k - \mathbf{G}\mathbf{Q}_k^\top\mathbf{W}_k$ where the group average is rotated before the scaling is applied but it is hard to see the justification and interpretation of weighting a group average after rotation. Many similar model variations may be envisaged but, for the reasons given above, we prefer the one we have adopted in (10).

In Layer 1 we used principal component analysis (PCA) to give an R-dimensional approximation to the Procrustean group average $\mathbf{G}$. This remains a possibility in Layer 2 but the component rotation would destroy the diagonal property of the saliences $\mathbf{W}_k$. When $p = 2$, component rotation is superfluous, supporting the use of initial two-dimensional individual scaling solutions $\mathbf{X}_k$, but in other cases we would have to reconcile using saliences in the group-average space in combination with a rotated group average. Perhaps this is not as bad as it may seem, because it is difficult to visualise three or more dimensional spaces in any set of circumstances, including INDSCAL. However, another way of dealing with the difficulty is to replace $\mathbf{Q}_k$ by projection matrices $\mathbf{P}_k$ as is described in Section 4.

The previous paragraph may be put into algebraic form as follows. Firstly we write the Procrustean criterion (Layer 1) as

$$\sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{Q}_k||^2 = \sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}||^2 + K||\mathbf{G}||^2 \ . \tag{11}$$

Then we express $\mathbf{G}$ in R dimensions as $\mathbf{G} = \mathbf{G}(\mathbf{I} - \mathbf{P}\mathbf{P}^\top) + \mathbf{G}(\mathbf{P}\mathbf{P}^\top)$, where $\mathbf{P}$ is an orthonormal matrix with R columns corresponding to the R leading eigenvalues of $\mathbf{G}^\top\mathbf{G}$. It

is the rows of $\mathbf{G}(\mathbf{P}\mathbf{P}^\top)$ or, equivalently in R dimensions, $\mathbf{GP}$ that give the R-dimensional principal components approximation for plotting the group average. With Layer 2, we may still use $\mathbf{GP}$ to give a principal components approximation to $\mathbf{G}$ but the question is what to do about $\mathbf{W}_k$? We could base an R-dimensional plot on $\mathbf{GPP}^\top\mathbf{W}_k$ giving $\mathbf{GP}$ as before and $\mathbf{W}_k\mathbf{P}$ as the saliences. The problem with this is that the transformed saliences are not diagonal and then it is hard to compare the R-dimensional saliences across the K groups.

Therefore, we try replacing the orthogonal matrices $\mathbf{Q}_k$ by R-dimensional orthonormal matrices $\mathbf{P}_k$ to give

$$\sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{P}_k||^2 = \sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{P}_k - \mathbf{G}||^2 + K||\mathbf{G}||^2 \ . \tag{12}$$

The orthogonal analysis of variance (11) implies that

$$\min_{\mathbf{Q}_k} \sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}||^2 \quad \text{and} \quad \max_{\mathbf{Q}_k} ||\mathbf{G}||^2$$

are attained for the same $\mathbf{Q}_k$ $(k = 1, \ldots, K)$. That is, for Procrustes INDSCAL with orthogonal matrices $\mathbf{Q}_k$, the minimal within variance occurs for the same orthogonal matrices as for the maximal group variance $\mathbf{G} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}_k\mathbf{Q}_k$ (see (8)). Unfortunately, this is not true for orthonormal matrices. Indeed,

$$\min_{\mathbf{P}_k} \sum_{k=1}^{K} ||\mathbf{X}_k\mathbf{P}_k - \mathbf{G}||^2 \quad \text{and} \quad \max_{\mathbf{P}_k} ||\mathbf{G}||^2$$

occur for different settings of $\mathbf{P}_k (k = 1, \ldots, K)$. However, although we have been interpreting our criteria in terms of least squares, because of the othogonality in (11) it is equally valid to interpret them in terms of maximising the group average $||\mathbf{G}||^2$. This interpretation is attractive and arguably better than least squares. Moreover it remains valid for projection matrices (cf. Section 4). Therefore we also include the criterion of maximising $||\mathbf{G}||^2$, where $\mathbf{G} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}_k\mathbf{P}_k$. Note that reduction to R dimensions is built into the criterion, so that there is no need for PCA. Also, there is no need to include $\mathbf{Q}_k$, to accommodate arbitrary rotations of the configurations, because $\mathbf{Q}_k\mathbf{P}_k$ is itself an orthonormal matrix.

Note that the Procrustean analogue of INDSCAL differs from the well-known *P*rocrustean *IN*dividual *DI*fferences *S*caling (PINDIS) family (Lingoes and Borg, 1978) in the sense that in the latter one always assumes that the group average is given or derived externally from, say, generalized Procrustes analysis. In the Procrustes problem (10), $\mathbf{G}$ is a matrix of model unknowns. Not only does (10) provide estimates of the group average $\mathbf{G}$ together with dimension weights $\mathbf{W}_k$, as with INDSCAL, but also individual configurations $\mathbf{X}_k\mathbf{Q}_k$ rotated into the position where they best match a scaled group average. As with INDSCAL itself, the estimated group average may not be rotated; indeed optimal rotations are included in the model and incorporated into the final estimates.

Third layer (Model $\mathbf{G}\mathbf{W}_k$, $\mathbf{G}$ orthonormal)

Finally, the orthonormal INDSCAL model imposes orthogonality constraints on the group average matrix in the form $\mathbf{G}^\top\mathbf{G} = \mathbf{I}_R$. This is a severe substantive constraint and the resulting model is substantively modified as was described in Section 1. Indeed, if the proximities were covariance or correlation matrices, then the orthonormal INDSCAL model may be identified as one of the variants of common principal component analysis (Flury, 1988; Pham, 2000), where the columns of $\mathbf{G}$ are the common eigenvectors of $\mathbf{B}_1, \ldots, \mathbf{B}_K$ and $\mathbf{W}_1^2, \ldots, \mathbf{W}_K^2$ are the corresponding eigenvalues that vary from group to group.

However, in its configurational version, differences exist between common components analysis and the Procrustean methods (Gower and Dijksterhuis, 2004, pp. 181-182). In Procrustes analysis, the rows do match, that is, the cases are common to all the K configurations. In contrast, the columns of $\mathbf{X}_k$ do not match. Then, the configuration version of orthornormal INDSCAL aims to match the K configurations to the best rank R configuration, that is, we minimize $\sum_{k=1}^{K} ||\mathbf{X}_k - \hat{\mathbf{X}}||^2$, for specified $\hat{\mathbf{X}}$ of rank R. In common components analysis the emphasis is on variables, rather than configurations, coupled with no requirement for row-matching. Common components analysis tries to find principal components of data matrices $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$ that share the same variables. This suggests to us that one could carry out common components analysis directly on the data matrices, where the $\mathbf{X}_k$ in the Procrustean criterion (10) are replaced by $\mathbf{Y}_k$ $(k = 1, \ldots, K)$.

3 The orthogonal Procrustean analogue of INDSCAL

The (orthogonal) Procrustean analogue of INDSCAL is based on $n \times R$ configuration matrices $\mathbf{X}_k$ $(k = 1, \ldots, K)$ and minimizes the criterion (Gower and Dijksterhuis, 2004, p. 171):

$$g(\mathbf{G}, \mathbf{Q}_1, \ldots, \mathbf{Q}_K, \mathbf{W}_1, \ldots, \mathbf{W}_K) = \sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{Q}_k - \mathbf{G}\mathbf{W}_k\|^2 \ , \tag{13}$$

where $\mathbf{Q}_k^\top\mathbf{Q}_k = \mathbf{Q}_k\mathbf{Q}_k^\top = \mathbf{I}_R$ for $k = 1, \ldots, K$, $\mathbf{G}$ denotes an $n \times R$ matrix and $\mathbf{W}_1, \ldots, \mathbf{W}_K$ are $R \times R$ diagonal matrices. The aim is to minimize the objective function (13) subject to one of the identification constraints (3) or (4). The configuration matrices $\mathbf{X}_k$ may be given, but, more likely, they will have been derived using some form of MDS. One possibility is to use classical scaling but for robustness, one may prefer to use other MDS methods such as least squares scaling by minimizing STRESS.

3.1 Finding the group average

Making use of the identification constraint (3), expanding (13) gives:

$$g \ = \ \sum_{k=1}^{K} ||\mathbf{X}_k||^2 + K\text{trace}\mathbf{G}\mathbf{G}^\top - 2\text{trace}\left(\sum_{k=1}^{K} \mathbf{X}_k\mathbf{Q}_k\mathbf{W}_k\right)\mathbf{G}^\top. \tag{14}$$

Differentiation of g with respect to $\mathbf{G}$ gives:

$$\mathbf{G} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{X}_k\mathbf{Q}_k\mathbf{W}_k \ . \tag{15}$$

This is a fundamental result, that shows that the estimated $\mathbf{G}$ is a simple group average of the individual rotated and weighted configurations. Note that this is not a condition imposed by the criterion (14) but is a consequence. The orthogonal rotations $\mathbf{X}_k\mathbf{Q}_k$ conform with the inner-products and Euclidean distances of the initial configurations, while the weights are unique and are not emenable to further rotation as they would be in a convential Procrustes analysis. These properties parallel the usual geometrical visualisations and interpretations both of Generalised Procrustes analysis and of INDSCAL itself. Further, (14) and (15) satisfy the orthogonal analysis of variance (ANOVA): $\sum_{k=1}^{K} ||\mathbf{X}_k||^2 = g + K\text{trace}(\mathbf{G}\mathbf{G}^\top)$. Note that minimising g is the same as maximising the group average $||\mathbf{G}||^2$ (cf. Section 4).

3.2 ALS algorithm

We now consider an ALS algorithm for estimating $\mathbf{G}$, $\mathbf{Q}_k$ and $\mathbf{W}_k$. Equation (15) shows that $\mathbf{G}$ is easily estimated from the current values of $\mathbf{Q}_k$ and $\mathbf{W}_k$. The rotations $\mathbf{Q}_k$ that minimize (14), keeping the unknowns $\mathbf{G}$ and $\mathbf{W}_k$ fixed, are estimated as $\mathbf{Q}_k = \mathbf{V}_k\mathbf{U}_k^\top$, where $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top$ is the SVD of $\mathbf{W}_k\mathbf{G}^\top\mathbf{X}_k$.

In common with Generalised Procrustes Analysis with isotropic scaling, there is an option for estimating anisotropic scales $\mathbf{W}_k$ either as an eigenvalue problem or incrementally. Gower and Dijksterhuis (2004) present a step to find an estimate of $\mathbf{W}_k$ that involves solving R eigenvalue problems. Here we give the details of an incremental alternative. Inserting the group average (15) into (14) gives:

$$\begin{aligned} g &= \sum_{k=1}^{K} ||\mathbf{X}_k||^2 - \text{trace}\left(\sum_{k=1}^{K} \mathbf{X}_k\mathbf{Q}_k\mathbf{W}_k\right)\mathbf{G}^\top \\ &= \sum_{k=1}^{K} ||\mathbf{X}_k||^2 - \sum_{k=1}^{K} \text{trace}[\text{diag}(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k)\mathbf{W}_k] \ . \end{aligned} \tag{16}$$

Estimating $\mathbf{W}_k$ requires R Lagrange multipliers associated with the R diagonal elements of the constraint (3). We imagine these to be collected in the diagonal of a matrix $\mathbf{L}$. Differentiating the Lagrangian g_L:

$$g_L = g + \text{trace}\left(\mathbf{L}\left(\sum_{k=1}^{K}\mathbf{W}_k^2 - K\mathbf{I}_R\right)\right) \tag{17}$$

with respect to $\mathbf{W}_k$ gives

$$\mathbf{W}_k = \mathbf{L}^{-1}\text{diag}(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k) \ . \tag{18}$$

Inserting (3) into (18) gives:

$$\sum_{k=1}^{K}\text{diag}^2(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k) = \mathbf{L}^2\sum_{k=1}^{K}\mathbf{W}_k^2 = K\mathbf{L}^2 \ . \tag{19}$$

Thus, the required matrices of weights $\mathbf{W}_k$ are given by

$$\mathbf{W}_k = \mathbf{L}^{-1}\text{diag}(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k) = \sqrt{K}\,\text{diag}(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k)\left(\sum_{k=1}^{K}\text{diag}^2(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k)\right)^{-1/2} \ , \tag{20}$$

provided that the diagonal matrix $\sum_{k=1}^{K}\text{diag}^2(\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k)$ is not singular.

One could simply start the algorithm with all $\mathbf{W}_k = \mathbf{Q}_k = \mathbf{I}_R$. This amounts to starting $\mathbf{G}$

as the average of the $\mathbf{X}_k$. The first step is then to transform the current values of $\mathbf{X}_k\mathbf{Q}_k$ to fit $\mathbf{G}\mathbf{W}_k$. Then with the new $\mathbf{X}_k\mathbf{Q}_k$ evaluate $\mathbf{G}^\top\mathbf{X}_k\mathbf{Q}_k$ and then from (20) derive $\mathbf{W}_k$. Finally, recompute $\mathbf{G}$. Consolidating these results, suggests an ALS algorithm incorporating the following steps:

1. Set convergence criterion ϵ to some small value, say 10^{-6}. Initialize $\mathbf{W}_k$, $\mathbf{Q}_k$ ($k = 1, \ldots, K$) and $\mathbf{G}$ as $\mathbf{W}_k^{\mathrm{c}}$, $\mathbf{Q}_k^{\mathrm{c}}$ and $\mathbf{G}^{\mathrm{c}}$, respectively. Set the iteration counter $c = 0$.

2. Compute $g^c = g(\mathbf{G}^{\mathrm{c}}, \mathbf{Q}_1^{\mathrm{c}}, \ldots, \mathbf{Q}_K^{\mathrm{c}}, \mathbf{W}_1^{\mathrm{c}}, \ldots, \mathbf{W}_K^{\mathrm{c}})$.

3. For fixed $\mathbf{W}_k^{\mathrm{c}}$ and $\mathbf{G}^{\mathrm{c}}$, evaluate $\mathrm{SVD}(\mathbf{W}_k^{\mathrm{c}}\mathbf{G}^{\mathrm{c}^\top}\mathbf{X}_k) = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top$ and set $\mathbf{Q}_k^{\mathrm{c}+1} = \mathbf{V}_k\mathbf{U}_k^\top$ for $k = 1, \ldots, K$.

4. For fixed $\mathbf{Q}_k^{\mathrm{c}+1}$ and $\mathbf{G}^{\mathrm{c}}$, evaluate $\mathbf{G}^{\mathrm{c}^\top}\mathbf{X}_k\mathbf{Q}_k^{\mathrm{c}+1}$ and (20) to find an update $\mathbf{W}_k^{\mathrm{c}+1}$ for $k = 1, \ldots, K$.

5. For fixed $\mathbf{Q}_k^{\mathrm{c}+1}$ and $\mathbf{W}_k^{\mathrm{c}+1}$, set $\mathbf{G}^{\mathrm{c}+1} = \frac{1}{K}\sum_{k=1}^{K}(\mathbf{X}_k\mathbf{Q}_k^{\mathrm{c}+1}\mathbf{W}_k^{\mathrm{c}+1})$.

6. Compute $g^{c+1} = g(\mathbf{G}^{\mathrm{c}+1}, \mathbf{Q}_1^{\mathrm{c}+1}, \ldots, \mathbf{Q}_K^{\mathrm{c}+1}, \mathbf{W}_1^{\mathrm{c}+1}, \ldots, \mathbf{W}_K^{\mathrm{c}+1})$.

7. If $|g^c - g^{c+1}| > \epsilon$, set $c = c+1$ and go to step 3; else consider the algorithm converged.

Experience with Generalised Procrustes Analysis suggests that there are a myriad of minor algorithmic variants that might be investigated when we include isotropic scaling. We have already mention the choice between incremental and eigenvalue estimates of the weighting matrices. Another possibility is, at appropriate points, to use k-excluded group averages (i.e. where the k-th group is excluded from the averaging process - see Gower and Dijksterhuis (2004)). Yet another possibility is to choose to update $\mathbf{G}$ whenever $\mathbf{W}_k$ and/or $\mathbf{Q}_k$ are changed, or to leave the updating to the end of a cycle. Various advantages have been claimed for all variants but the practical differences seem slight; we would expect this to continue to be the case when anisotropic scaling is used.

4 A configurational version of INDSCAL with projection matrices

In the previous section we pointed out that although $n \times R$ configuration matrices $\mathbf{X}_k$ may be given, more likely, they will have been derived from $n \times p$ data matrices $\mathbf{Y}_k$ (with $p > R$) by using some form of MDS. With classical scaling, big matrices will have been orthogonally projected onto fewer dimensions and a similar effect is gained by using other methods of MDS. An advantage of using MDS is that the resulting configurations $\mathbf{X}_k$ will be small and so will the orthogonal matrices $\mathbf{Q}_k$ of Section 4.

However, we may wish to operate directly on the bigger matrices $\mathbf{Y}_k$, as when $n \times p$ raw data matrices $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$ are available with all p variables in the K sets being measured in the same units, so then $\mathbf{X}_k = \mathbf{Y}_k$ $(k = 1, \ldots, K)$ with $R = p$. In this case, minimizing (14) may lead to a high-dimensional optimization problem with large orthogonal matrices $\mathbf{Q}_k$. One is tempted to solve the following Procrustes problem instead:

$$\min \sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{P}_k - \mathbf{G}\mathbf{W}_k\|^2 \ , \tag{21}$$

where $\mathbf{P}_k$ $(k = 1, \ldots, K)$ are $p \times R$ column-wise orthonormal matrices. The matrices $\mathbf{P}_k$ are often called projection matrices (Gower and Dijksterhuis, 2004), because they have an interpretation of rotating an orthogonal projection of the p-dimensional configuration $\mathbf{X}_k$ to match an R-dimensional configuration $\mathbf{G}\mathbf{W}_k$.

Nevertheless, the attractiveness of the similarity of the criterion in (21) to (13) is deceptive. This can be seen by considering the matrices $\mathbf{X}_k = [\mathbf{A}\ \mathbf{B}_k]$ $(k = 1, \ldots, K)$, that is, $\mathbf{X}_k$ is partitioned into an $n \times R$ matrix $\mathbf{A}$ and an $n \times (p-R)$ matrix $\mathbf{B}_k$. For this example, setting $\mathbf{P}_k = \begin{pmatrix} \mathbf{I}_R \\ \mathbf{O}_{p-R} \end{pmatrix}$ gives $\mathbf{X}_k\mathbf{P}_k = \mathbf{A}$ and $\mathbf{W}_k = \mathbf{I}_R$ for $k = 1, \ldots, K$. Then $\mathbf{G} = \mathbf{A}$ and the residual sum of squares of the criterion in (21) is zero. The fit is perfect even though the projected configurations may have very bad residual fits $\mathbf{B}_1, \ldots, \mathbf{B}_k$ to the initial configurations. This may be an artificial example but it is not a pathological example, except in that it might be described as being pathologically well-behaved. In general (21) aims to minimise the sum of squares of the projected error and this seems not compatible to giving good approximations of the individual scalings in INDSCAL. As pointed out

at the end of Section 3, minimising the sum-of-squares (13) is equivalent to maximising the group average; this equivalence does not carry over to the projection criterion (21), because the orthogonal analysis of variance is no longer available. Instead, we propose to maximise the projected group average configuration, following the suggestion of Peay (1988).

4.1 Maximising the group average configuration

Consider maximising the following objective function:

$$h(\mathbf{P}_1, \ldots, \mathbf{P}_K, \mathbf{W}_1, \ldots, \mathbf{W}_K) = \left\| \sum_{k=1}^{K} \mathbf{X}_k \mathbf{P}_k \mathbf{W}_k \right\|^2 \quad , \tag{22}$$

subject to $\mathbf{P}_k^\top \mathbf{P}_k = \mathbf{I}_R$ $(k = 1, \ldots, K)$ and the identification constraint $\sum_{k=1}^{K} \mathbf{W}_k^2 = K\mathbf{I}_R$. Maximising (22) amounts to maximising the squared sum of the elements of the group average $K\mathbf{G} = \sum_{k=1}^{K} \mathbf{X}_k \mathbf{P}_k \mathbf{W}_k$. Searching for the maximal group average configuration was introduced by Peay (1988) in projection Procrustes analysis. Subsection 4.2 describes an ALS algorithm for maximising (22).

4.2 Algorithm

Expanding (22) gives:

$$\begin{aligned} h = K^2||\mathbf{G}||^2 &= K^2 \text{trace}\left(\mathbf{G}\mathbf{G}^\top\right) \\ &= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \text{trace}\left(\mathbf{X}_{k_1}\mathbf{P}_{k_1}\mathbf{W}_{k_1}\mathbf{W}_{k_2}\mathbf{P}_{k_2}^\top\mathbf{X}_{k_2}^\top\right) \\ &= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \text{trace}[\mathbf{W}_{k_1}\mathbf{W}_{k_2}\text{diag}(\mathbf{P}_{k_2}^\top\mathbf{X}_{k_2}^\top\mathbf{X}_{k_1}\mathbf{P}_{k_1})] \\ &= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} v(\mathbf{W}_{k_1})^\top[v(\mathbf{P}_{k_2}^\top\mathbf{X}_{k_2}^\top\mathbf{X}_{k_1}\mathbf{P}_{k_1}) \odot v(\mathbf{W}_{k_2})] \ , \end{aligned} \tag{23}$$

where $v(\mathbf{A})$ denotes the vector containing the main diagonal of the square matrix $\mathbf{A}$, i.e. if $\mathbf{A}$ is $n \times n$ then $v(\mathbf{A}) = (\mathbf{A} \odot \mathbf{I}_n)\mathbf{1}_n$ with $\mathbf{1}_n$ being a column vector of n ones.

Estimating $\mathbf{W}_k$ requires R Lagrange multipliers associated with the R diagonal elements of the constraint (3). We imagine these to be collected in the diagonal of a matrix $\mathbf{L}$.

Differentiating the Lagrangian h_L:

$$h_L = h + \text{trace}\left(\mathbf{L}\left(\sum_{k=1}^{K}\mathbf{W}_k^2 - K\mathbf{I}_R\right)\right) \tag{24}$$

with respect to $\mathbf{W}_k$ gives

$$\sum_{l=1}^{K}\text{diag}(\mathbf{P}_k^\top\mathbf{X}_k^\top\mathbf{X}_l\mathbf{P}_l)\mathbf{W}_l v(\mathbf{\Sigma}^{(kl)}) \odot v(\mathbf{W}_j) = \mathbf{L}v(\mathbf{W}_k) \ , \tag{25}$$

where $\mathbf{\Sigma}^{(kl)} = \mathbf{P}_k^\top\mathbf{X}_k^\top\mathbf{X}_l\mathbf{P}_l$. Equation (25) can be rewritten as

$$\sum_{l=1}^{K} v(\mathbf{\Sigma}^{(kl)}) \odot v(\mathbf{W}_l) = v(\mathbf{W}_k) \odot v(\mathbf{L}) \ . \tag{26}$$

Note that $v(\mathbf{\Sigma}^{(kl)}) = [\mathbf{X}_k\mathbf{P}_k \odot \mathbf{X}_l\mathbf{P}_l]^\top\mathbf{1}_n$, that is, there is no need to calculate the matrix $\mathbf{P}_k^\top\mathbf{X}_k^\top\mathbf{X}_l\mathbf{P}_l$. Let $\sigma_r^{(kl)}$ denote the rth element of $v(\mathbf{\Sigma}^{(kl)})$, and $w_r^{(k)}$ and l_r be the rth elements of $\mathbf{W}_k$ and $\mathbf{L}$ $(r = 1, \ldots, R)$, respectively. Then, for each Lagrange multiplier l_r the optimality condition (26) implies the following eigenvalue problem:

$$\begin{pmatrix} \sigma_r^{(11)} & \sigma_r^{(12)} & \ldots & \ldots & \sigma_r^{(1K)} \\ \sigma_r^{(21)} & \sigma_r^{(22)} & \ldots & \ldots & \sigma_r^{(2K)} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \sigma_r^{(K1)} & \sigma_r^{(K2)} & \ldots & \ldots & \sigma_r^{(KK)} \end{pmatrix} \begin{pmatrix} w_r^{(1)} \\ w_r^{(2)} \\ \vdots \\ \vdots \\ w_r^{(K)} \end{pmatrix} = l_r \begin{pmatrix} w_r^{(1)} \\ w_r^{(2)} \\ \vdots \\ \vdots \\ w_r^{(K)} \end{pmatrix} \ . \tag{27}$$

Thus, the K weight matrices $\mathbf{W}_k$ are found by solving R eigenvalue problems defined in (27). Note, that in most applications $R = 2$ or $R = 3$. If the normalized eigenvectors of (27) form a $K \times R$ matrix, then the kth row of this matrix gives the diagonal elements of $\mathbf{W}_k$. Instead of solving R eigenvalue problems, one may modify the incremental method described in the previous Section to find the weight matrices $\mathbf{W}_k$ $(k = 1, \ldots, K)$.

To find an update for $\mathbf{P}_k$ $(k = 1, \ldots, K)$, consider our Projection INDSCAL criterion (22) rewritten as

$$h = ||\mathbf{X}_k\mathbf{P}_k\mathbf{W}_k + (K-1)\mathbf{G}_k||^2 \ , \tag{28}$$

where $\mathbf{G}_k = 1/(K-1)\sum_{i\neq k}^{K}\mathbf{X}_i\mathbf{P}_i\mathbf{W}_i$ is the k-excluded group average, which gives the mean of all the current $\mathbf{X}_i\mathbf{P}_i\mathbf{W}_i$ excluding $\mathbf{X}_k\mathbf{P}_k\mathbf{W}_k$. Thus, for fixed $\mathbf{W}_k$ $(k = 1 \ldots, K)$,

the problem of maximising (22) is transformed into the two-set problem of maximising (28) for which all matrix parameters are known except for the column-wise orthonormal projection matrices $\mathbf{P}_k$ $(k = 1, \ldots, K)$. One can update each $\mathbf{P}_k$ sequentially, based on an algorithm of Koschat and Swayne (1991) developed in a different context, as follows. We consider writing (28) in the form

$$||\mathbf{X}_k\mathbf{P}_k\mathbf{W}_k + (K-1)\mathbf{G}_k||^2 = \left|\left| \begin{pmatrix} \mathbf{X}_k\mathbf{P}_k \\ \mathbf{X}_0\mathbf{P}_k \end{pmatrix} \mathbf{W}_k - \begin{pmatrix} -(K-1)\mathbf{G}_k \\ \mathbf{X}_0\mathbf{P}_k\mathbf{W}_k \end{pmatrix} \right|\right|^2 \tag{29}$$

and choose $\mathbf{X}_0$ such that $\mathbf{X}_k^\top\mathbf{X}_k + \mathbf{X}_0^\top\mathbf{X}_0 = \rho^2\mathbf{I}_p$ so that $\mathbf{W}_k\mathbf{P}_k^\top(\mathbf{X}_k^\top\mathbf{X}_k + \mathbf{X}_0^\top\mathbf{X}_0)\mathbf{P}_k\mathbf{W}_k = \rho^2\mathbf{W}_k^2$ because $\mathbf{P}_k^\top\mathbf{P}_k = \mathbf{I}_R$. Then, in the $(c+1)$th iterative step for estimating $\mathbf{P}_k$, we have only to maximise trace$(\mathbf{W}_k[-(K-1)\mathbf{G}_k^\top\mathbf{X}_k + \mathbf{W}_k\mathbf{P}_k^{\mathrm{c}\top}\mathbf{X}_0^\top\mathbf{X}_0]\mathbf{P}_k^{\mathrm{c}+1\top})$, where $\mathbf{P}_k^{\mathrm{c}\top}$ is regarded as fixed. Thus, we form

$$\mathbf{Z} = \mathbf{W}_k[-(K-1)\mathbf{G}_k^\top\mathbf{X}_k + \mathbf{W}_k\mathbf{P}_k^{\mathrm{c}\top}(\rho^2\mathbf{I}_p - \mathbf{X}_k^\top\mathbf{X}_k)] \tag{30}$$

and then from the SVD $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ derive $\mathbf{P}_k^{\mathrm{c}+1} = \mathbf{V}\mathbf{U}^\top$ from $\mathbf{U}$ and the first R columns of $\mathbf{V}$, which correspond to the R non-zero singular values of the $R \times p$ matrix $\mathbf{Z}$. One may choose $\rho^2 = ||\mathbf{X}_k||^2$ or $\rho^2 = \lambda$, the maximal eigenvalue of $\mathbf{X}_k^\top\mathbf{X}_k$ (Gower and Dijksterhuis, 2004). Thus, we iterate on $\mathbf{P}_k^{\mathrm{c}}$ until convergence; proof of convergence of a similar algorithm is given in Gower and Dijksterhuis (2004). This procedure is carried out for all $\mathbf{P}_k$ $(k = 1, \ldots, K)$. Then, with the new $\mathbf{P}_k$ one finds new matrices $\mathbf{W}_k$. This alternating procedure is repeated until a pre-specified convergence criterion is met. Finally, once the algorithm has converged one can compute $\mathbf{G}$. Consolidating these results, suggests an ALS algorithm incorporating the following steps:

1. Set convergence criterion ϵ to some small value, say 10^{-6}. Initialize $\mathbf{W}_k$ and $\mathbf{P}_k$ $(k = 1, \ldots, K)$ as $\mathbf{W}_k^{\mathrm{c}}$ and $\mathbf{P}_k^{\mathrm{c}}$, respectively. Set the iteration counter $c = 0$.

2. Compute $h^c = h(\mathbf{P}_1^{\mathrm{c}}, \ldots, \mathbf{P}_K^{\mathrm{c}}, \mathbf{W}_1^{\mathrm{c}}, \ldots, \mathbf{W}_K^{\mathrm{c}})$.

3. For fixed $\mathbf{P}_k^{\mathrm{c}}$ solve the R eigenvalue problems defined in (27) to find $\mathbf{W}_k^{\mathrm{c}+1}$ for $k = 1, \ldots, K$.

4. For fixed $\mathbf{W}_k^{\mathrm{c}+1}$ find an update $\mathbf{P}_k^{\mathrm{c}+1}$ for $k = 1, \ldots, K$ as follows:

(a) Set $\rho^2 = ||\mathbf{X}_k||^2$ and $\mathbf{P}_1 = \mathbf{P}_k^{\mathrm{c}}$.

(b) Form $\mathbf{Z} = \mathbf{W}_k[-(K-1)\mathbf{G}_k^\top \mathbf{X}_k + \mathbf{W}_k \mathbf{P}_1^\top (\rho^2 \mathbf{I}_p - \mathbf{X}_k^\top \mathbf{X}_k)]$.

(c) Compute the SVD of $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ and set $\mathbf{P} = \mathbf{V}\mathbf{U}^\top$.

(d) If $||\mathbf{P} - \mathbf{P}_1||^2$ is greater than some threshold set $\mathbf{P}_1 = \mathbf{P}$ and go to step (b). Otherwise set $\mathbf{P}_k^{\mathrm{c}+1} = \mathbf{P}$.

5. Compute $h^{c+1} = h(\mathbf{P}_1^{\mathrm{c}+1}, \ldots, \mathbf{P}_K^{\mathrm{c}+1}, \mathbf{W}_1^{\mathrm{c}+1}, \ldots, \mathbf{W}_K^{\mathrm{c}+1})$.

6. If $|h^c - h^{c+1}| > \epsilon$, set $c = c + 1$ and go to step 3; else consider the algorithm converged.

5 Application to food science data

5.1 Description of the data

In this Section, the methods developed in this paper are illustrated with an application from sensory evaluation, a scientific discipline that applies principles of experimental design and statistical analysis to the use of human senses (sight, smell, taste, touch and hearing) for the purposes of evaluating consumer products. Sensory analysis requires panels of human assessors, on whom certain products are tested, and recording the responses made by them. By applying statistical techniques to the results it is possible to make inferences and insights about the products under test. Here, a sensory panel of tasters assesses several properties of food items. The data arose from one sensory profiling session, the last of a range of seven. The seven sessions took place in an experiment to develop a sensory vocabulary to describe warmed-over flavour in meat patties. For details of this study, the reader is referred to Byrne et al. (2001).

A panel of $K = 8$ assessors scored their perceived intensities of 20 properties, such as fresh pork meat-like-flavour, roasted like odour or sweet taste, of $n = 6$ different pork meat patties. Five properties, which take constant values for some judges, were removed, hence $p = 15$. That is, eight (6×15) data matrices $\mathbf{Y}_k$ $(k = 1, \ldots, 8)$ are given with the six food items in its rows and the fifteen attributes in its columns. This data structure is a three-mode structure and it is presupposed that all the K assessors use the same p

attributes, and that these are presented in the same order for each $\mathbf{Y}_k$. An INDSCAL matrix decomposition model with $R = 2$ dimensions is considered in the following. Configuration matrices $\mathbf{X}_k \in \mathbb{R}^{6\times 2}$ are derived from distance matrices $\mathbf{D}_k$ $(k = 1, \ldots, 8)$ by means of metric MDS. These configuration matrices $\mathbf{X}_k$ are the data to which the orthogonal Procrustean analogue of INDSCAL was applied. For the configurational version of INDSCAL with projection matrices, the matrices $\mathbf{Y}_k = \mathbf{X}_k$ $(k = 1, \ldots, 8)$ of dimension 6×15 were taken as input data. For comparison, inner-product matrices, derived from the spectral decomposition $\mathbf{B}_k = \mathbf{X}_k\mathbf{X}_k^\top$, were analysed using the original INDSCAL method of Carroll and Chang (1970).

Computations were carried out using MATLAB Version 7.10.0 (R2010a) (The MathWorks, 2010) on a PC operating with an Intel Pentium 4 CPU having 2.4 GHz clock frequency and 1GB of RAM. All computer codes used in the experiments are available upon request.

5.2 Results

Figure 1 (upper graph) presents the group average configuration for the six food items (A, B, C, D, E, F) for both INDSCAL given by Carroll and Chang (1970) and the orthogonal Procrustean analogue of INDSCAL.

* * * Figure 1 about here * * *

Although the exact positions for the food items differ, the configurations are very similar. In INDSCAL, the relationships of the individual configurations to the group average are summarised by weights to be attached to the axes of the upper graph of Figure 1; these are shown in the lower graph of Figure 1. The largest weights are those of assessors 1 and 5, associated with the horizontal axis, and assessor 2 and 3, associated with the vertical axis of the INDSCAL group stimulus space. With respect to the second underlying dimension representing the attributes the two sets of results diverge most for assessors 3, 4, and 7. Whereas in the solution for the configurational version the subject weights for assessors 4 and 7 are much larger with respect to the second dimension, the INDSCAL solution obtained by Carroll and Chang (1970) shows larger weights for the same dimension for assessor 3. Inspections of the correlations of the individual attributes with the dimensions

of the group average for approaches revealed that the assessors differ in their use of certain attributes. Results for the configurational version with projection matrices (not shown) are very similar to the ones displayed in Figure 1 for the orthogonal Procrustean analogue of INDSCAL.

It is interesting to compare the goodness-of-fit obtained by the Procrustes version of INDSCAL with the one obtained by Carroll and Chang (1970). As the fitting criteria (2) and (13) are different, their minima are not directly comparable. Instead, the relative error of fit is used to compare the standard INDSCAL solution with the solution obtained by the orthogonal Procrustean analogue, where relative error of fit is defined as the minimum fit obtained for the corresponding problem divided by the Frobenius norm of the fitted data. Thus, for the two alternative methods of three-way individual differences scaling, the following discrepancy of fit measures are compared:

$$\sum_{k=1}^{K} \frac{||\mathbf{B}_k - \mathbf{G}\mathbf{W}_k^2\mathbf{G}^\top||^2}{||\mathbf{B}_k||^2} \tag{31}$$

for the inner-product-version and

$$\sum_{k=1}^{K} \frac{||\mathbf{X}_k - \mathbf{G}\mathbf{W}_k\mathbf{Q}_k^\top||^2}{||\mathbf{X}_k||^2} \tag{32}$$

for the orthogonal Procrustean analogue of INDSCAL. The value of the loss function (31) is 3.2082 and the value of the loss function (32) is 2.1721 (using rational starts for the unknown matrix parameters). The latter decreases to 0.0083 if one uses $\mathbf{X}_k = \mathbf{Y}_k$ $(k = 1, \ldots, K)$ as input matrices and applies the configuration version with column-wise projection matrices $\mathbf{P}_k$ instead of using $\mathbf{Q}_k$ $(k = 1, \ldots, K)$. This shows that the configuration versions give a better fit to the data than the inner-product version. This is not surprising as the Procrustean versions are formulated in terms of first-order matrices $\mathbf{X}_k$ rather than second-order matrices $\mathbf{B}_k$.

The model parameter matrices $\mathbf{Q}_k$ and $\mathbf{P}_k$ $(k = 1, \ldots, K)$ are initialized randomly, whereas for the weight matrices $\mathbf{W}_k = \mathbf{I}_R$ for $k = 1, \ldots, K$ are used. To avoid local optima, the algorithm was run twenty times and it was stopped when successive function values differed by less than $\epsilon = 10^{-4}$. For $R = 2$, the procedure for fitting the orthogonal Procrustean analogue of INDSCAL required on average 184 iterations to converge,

taking about 0.15 seconds. For the configuration version with column-wise orthonormal 15×2 matrices $\mathbf{P}_k$ $(k = 1 \ldots, K)$ the algorithm required on average 130 iterations, taking about 1.5 seconds. Recall that the configuration version of INDSCAL with projection matrices requires R eigenvalue calculations at each iteration for finding an update for $\mathbf{W}_k$ $(k = 1, \ldots, K)$. The twenty runs led to the same function value, up to the second decimal place, which was deemed adequate. Using a higher accuracy criterion such as $\epsilon = 10^{-6}$ needed considerably more CPU time but did not change the quality of the solution.
To give some insight into the iteration process, for two of the twenty randomly started runs the function value has been plotted against the iteration number in Figure 2 for the configuration version of INDSCAL with orthogonal matrices and column-wise orthonormal (projection) matrices, respectively.

* * * Figure 2 about here * * *

It can be seen that for both algorithms the decrease (increase) of the objective function is rather gradual. In the left (right) plot of Figure 2, the monotonically decreasing (increasing) function value stabilizes after 184 iterations (85 iterations) at a value of 804.7727 (194723.1113).

6 Discussion

In this paper, INDSCAL is considered as being embedded within a hierarchy of individual difference scaling models. We explore the members of this family, distinguishing models, the role of identification and substantive constraints, criteria for fitting models and algorithms to optimise the criteria. Our approach extends previously published methodology on INDSCAL through adding new model variants, which are based on data in the form of configurational matrices. We introduced configurational versions of INDSCAL based on orthogonal and column-wise orthonormal (projection) matrices, respectively. Algorithms are introduced for fitting the new models. We illustrated our methods with a typical application in sensory evaluation. For this particular example, the obtained group average and saliences are similar to what we obtained using the classic approach to INDSCAL of Carroll and Chang (1970), but the Procrustean versions resulted in a better fit to the

data than the inner-product version. In fact, the two sets of results, although similar, do show some quite large differences in the individual space.

There are three broad limitations to our approach. The first two limitations, which are shared with other methods of individual differences scaling, are due to issues inevitably associated with the INDSCAL model. While INDSCAL is based on a model, this model could be described as deterministic rather than stochastic, in the sense that no distributional assumptions are made about the error terms. Consequently, a drawback of the INDSCAL decomposition models with fixed matrix parameters is that it is not possible to test them by statistical methods. Nevertheless, the notions of convergence of the algorithms, stability of solutions, and goodness-of-fit continue to apply. Furthermore, INDSCAL assumes that the same fundamental dimensions govern all subjects' perceptions of the set of stimuli, with just the weight allocated to each dimension differing between subjects. Whether the INDSCAL assumption is plausible in analysing different people's perception of the similiarities between various consumer products such as food items, is certainly open to question. The third limitation is that the loss functions of the Procrustean versions of INDSCAL are different from those of the inner-product version or the ones that operate directly on distance matrices. Thus, if we are going to compare models, we should make sure that we are comparing like with like, and comparing models is a minefield that many authors seem prepared to cross. Therefore, we focused on comparing some common aspects of the two models, such as the group average solution or individual weights. We believe that statements based on such a comparison are reasonably valid.

Future work, some of it in process, is needed in two areas. Firstly, other fitting criteria for the configurational version of INDSCAL with projection matrices could be envisaged. For example, one could try to solve the following maximisation problem:

$$\max \frac{\left\| \sum_{k=1}^{K} \mathbf{X}_k \mathbf{P}_k \right\|^2}{\left\| \sum_{k=1}^{K} \mathbf{X}_k \mathbf{P}_k - \mathbf{G}\mathbf{W}_k \right\|^2} \quad \text{subject to} \quad \sum_{k=1}^{K} \mathbf{W}_k^2 = K\mathbf{I} \ . \tag{33}$$

The criterion in (33) models the individual differences as in the rest of the paper. It seeks to maximise the group average, but simultaneously seeks to minimise the differences

between each projected configuration and the scaled group average. Finding a suitable algorithm to maximise (33) may be a fruitful line of research. Secondly, the proposed algorithms in this paper have been of an alternating least squares type and more knowledge is needed about their performance with respect to speed, accuracy, and the ability to find global rather than local optima, applied to both real and artificial data. One may also compare these algorithms with different algorithms for optimising the same criterion. More work is required here.

Acknowledgements

We are grateful to Garmt B. Dijksterhuis who provided the food data and to Henk A. L. Kiers for sending us a MATLAB implementation of the original INDSCAL algorithm by Carroll and Chang (1970).

References

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008), *Optimization Algorithms on Matrix Manifolds*, Princeton University Press: Princeton.

Borg, I. and Groenen, P. J. F. (2005), *Modern Multidimensional Scaling: Theory and Applications*, Springer: New York, 2nd ed.

Byrne, D. V., O'Sullivan, M. G., Dijksterhuis, G. B., Bredie, W. L. P., and Martens, M. (2001), "Sensory panel consistency during developement of a vocabulary for warmed-over flavour," *Food Quality and Preference*, 12, 171–187.

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.

— (1972), "IDIOSCAL: A generalization of INDSCAL allowing IDIOsyncratic references systems as well as an analytic approximation to INDSCAL," .

Cox, T. F. and Cox, M. A. A. (2001), *Multidimensional Scaling*, Chapman & Hall: Boca Raton, Florida, 2nd ed.

De Leeuw, J. (1977), "Applications of convex analysis to multidimensional scaling," in *Recent Developments in Statistics*, eds. Barra, J. R., Brodeau, F., Romier, G., and van Cutsem, B., Amsterdam: North-Holland, pp. 133–145.

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

De Leeuw, J. and Mair, P. (2009), "Multidimensional scaling using majorization: SMACOF in R," *Journal of Statistical Software*, 31, 1–30.

Flury, B. (1988), *Common Principal Components and Related Multivariate Models*, John Wiley & Sons: New York, New York.

Gower, J. C. and Dijksterhuis, G. B. (2004), *Procrustes Problems*, Oxford University Press: Oxford.

Harshman, R. A. (1978), "Models for analysis of asymmetrical relationships among N objects or stimuli," Paper presented at the First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology, McMaster University, Hamilton, Ontario.

— (1982), "DEDICOM: a family of models generalizing factor analysis and multidimensional scaling for decomposition of asymmetric relationships," Unpublished manuscript, University of Western Ontario.

Kiers, H. A. L. (1989), "A computational short-cut for INDSCAL with orthonormality constraints on positive semi-definite matrices of low rank," *Computational Statistics Quarterly*, 2, 119–135.

Koschat, M. A. and Swayne, D. F. (1991), "A weighted Procrustes criterion," *Psychometrika*, 56, 229–239.

Krijnen, W. P., Dijkstra, T. K., and Stegeman, A. (2008), "On the non-existence of optimal solutions and the occurence of "degeneracy" in the CANDECOMP/PARAFAC model," *Psychometrika*, 73, 431–439.

Kroonenberg, P. M. (1983), *Three-Mode Principal Component Analysis: Theory and Applications*, DSWO Press: Leiden.

— (2008), *Applied Multiway Data Analysis*, John Wiley & Sons: Hoboken, New Jersey.

Lingoes, J. C. and Borg, I. (1978), "A direct approach to individual differences scaling using increasingly complex transformations," *Psychometrika*, 43, 491–519.

Peay, E. R. (1988), "Multidimensional rotation and scaling of configurations to optimal agreement," *Psychometrika*, 53, 199–208.

Pham, D. T. (2000), "Joint approximate diagonalization of positive definite hermitian matrices," *SIAM Journal on Matrix Analysis and Applications*, 22, 1136–1152.

Stegeman, A. (2007), "Degeneracy in CANDECOMP/PARAFAC explained for $p \times p \times 2$ arrays of rank $p + 1$ or higher," *Psychometrika*, 71, 483–501.

Takane, Y., Jung, K., and Hwang, H. (2010), "An acceleration method for Ten Berge et al.'s algorithm for orthogonal INDSCAL," *Computational Statistics*, 25, 409–428.

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

ten Berge, J. M. F. and Kiers, H. A. L. (1991), "Some clarifications of the CANDECOMP algorithm applied to INDSCAL," *Psychometrika*, 56, 317–326.

ten Berge, J. M. F., Kiers, H. A. L., and Krijnen, W. P. (1993), "Computational solutions for the problem of negative saliences and nonsymmetry in INDSCAL," *Journal of Classification*, 10, 115–124.

ten Berge, J. M. F., Knol, D. L., and Kiers, H. A. L. (1988), "A treatment of the orthomax rotation family in terms of diagonalization, and a re-examination of a singular value approach to varimax rotation," *Computational Statistics Quarterly*, 3, 207–217.

The MathWorks (2010), *MATLAB, Version 7.10.0 (R2010a)*, The MathWorks Inc.: Natick, Massachusetts.

Torgerson, W. S. (1958), *Theory and Methods of Scaling*, Wiley: New York.

Trendafilov, N. T. (2004), "Orthonormality-contrained INDSCAL with nonnegative saliences," in *Computational science and its applications - ICCSA 2004*, ed. Laganá, A. e. a., Springer: Berlin, Heidelberg, pp. 952–960.

— (2012), "DINDSCAL: direct INDSCAL," *Statistics and Computing*, 22, 445–454.

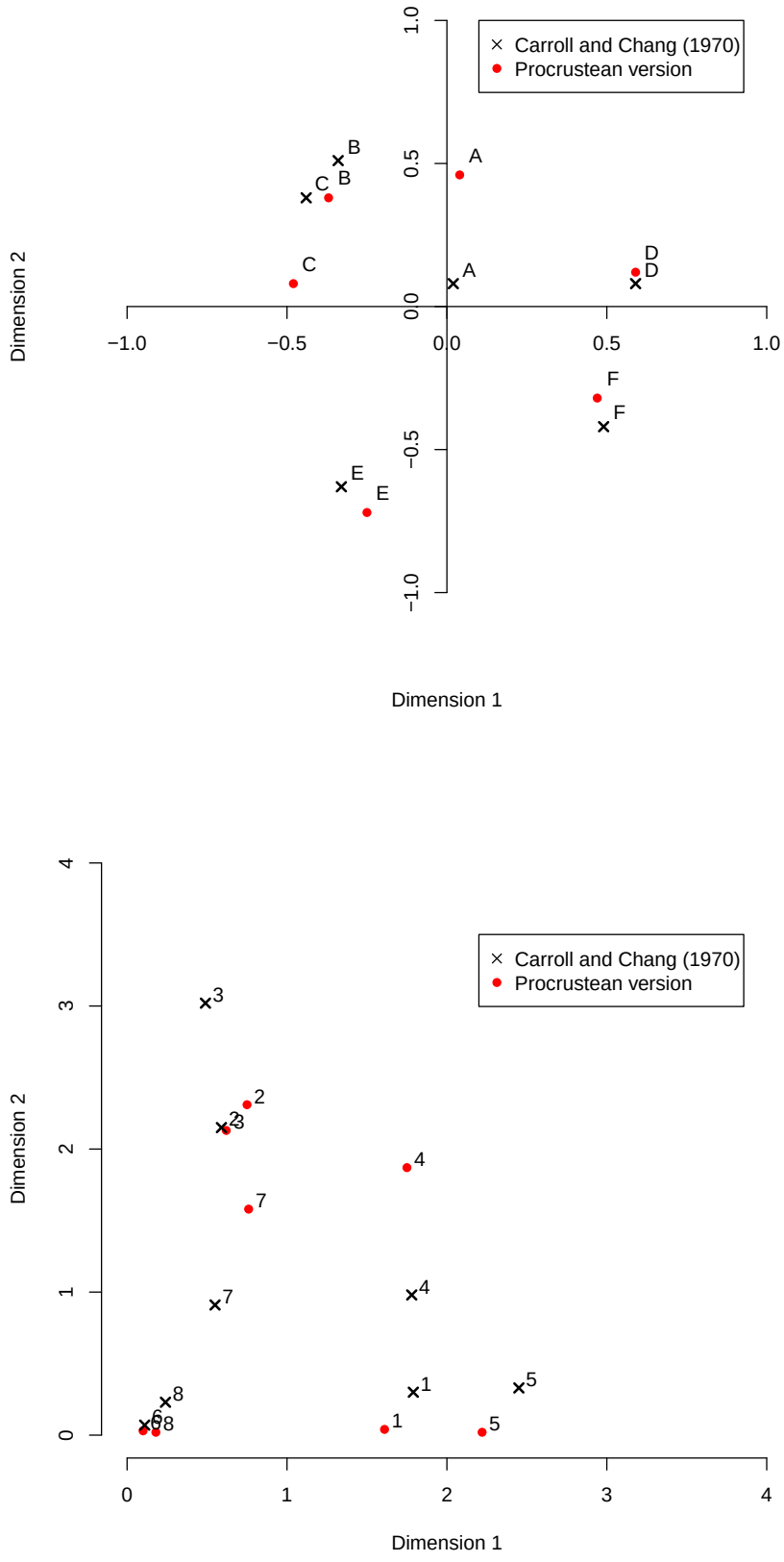

Figure 1: Group average of the six pork meat patties (upper graph) and weights of the eight assessors (lower graph), obtained by two variants of INDSCAL.

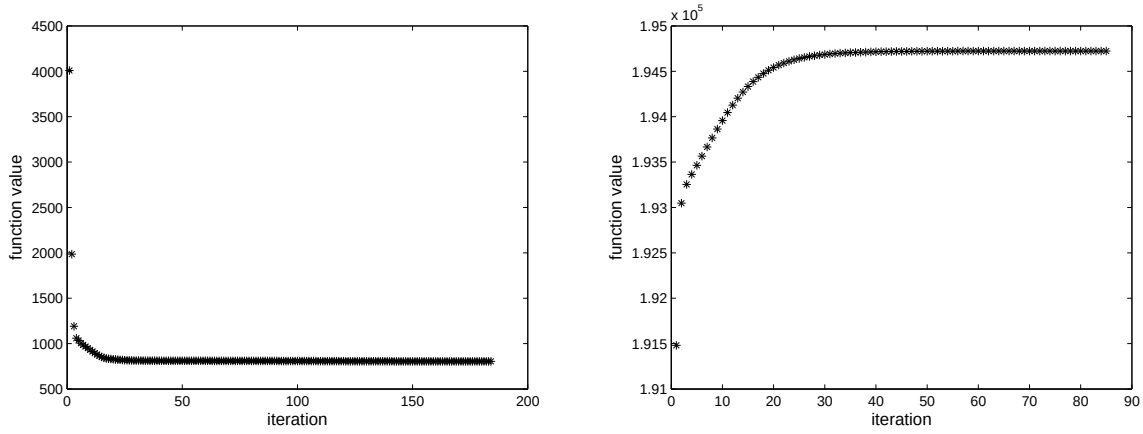

Figure 2: Function value plotted against iteration number for the configuration version of INDSCAL with orthogonal matrices (left) and column-wise orthonormal (projection) matrices (right), for two randomly started runs of the corresponding algorithm