RR-98-03

A New Approach to the Subpsace Matching of Configurations, Including Object Weights

Jacques J.F. Commandeur

Data Theory Group

Department of Education

Faculty of Social and Behavioral Sciences

Leiden University

This study is funded by The Netherlands Organization for Scientific Research (NWO) by grant 575-67-053 for the 'Pioneer' project 'Subject Oriented Multivariate Analysis'. The author is grateful to Wim Vaessen of Essensor in Delft for kindly making available the milk consumer research data.

Abstract

A new approach is proposed for the least-squares or Procrustes matching of two or more configurations containing different column orders. More generally, this least-squares criterion, which is not sensitive to "rotating out variance", can be used to obtain a solution in any subspace of the configurations whether they have equal or different column orders. The criterion is applied to three different models for the matching of configurations. The first model is Gower's generalized Procrustes analysis which consists of the matching of configurations under (relative) Euclidean distance preserving transformations. The other two are models proposed by Lingoes and Borg; they allow for differential weighting of dimensions, and are analogs of the INDSCAL and IDIOSCAL models for the analysis of (dis)similarity data. Convergent algorithms are provided for all three models. At the same time, the possibility of attaching weights to the rows of the configurations is incorporated in all algorithms. The handling of configurations with unequal row orders, then, is a matter of assigning zero weights to missing rows of the configurations.

Key words: matching, upward Procrustes rotation, object weights, individual differences analysis

1. Introduction

Multidimensional scaling as well as a number of multivariate analysis techniques can be used to obtain a spatial representation of objects, consisting of a geometric configuration of points, usually in a Euclidean space of limited dimensionality. However, configurations may also consist of ratings of objects on a number of variables. Matching procedures are concerned with the comparison of M ($M \geq 2$) configurations $\mathbf{X}_j$ ($j = 1, \ldots, M$) of order $N \times m_j$, where each configuration can be considered as containing the coordinates of the same N objects in m_j dimensions. Typically, in matching procedures the amount of agreement between configurations is studied under certain types of admissible transformations of the configurations.

In Generalized Procrustes Analysis (GPA) as developed by Gower (1975) the type of transformations is restricted to relative Euclidean distance preserving transformations, that is, translations, orthonormal transformations, and isotropic scaling. Using previous results of Green and Gower (1979), and ten Berge and Knol (1984), Peay (1988) discussed how to apply all these transformations in the case where the configurations have different column orders as well as how to obtain an optimal GPA solution in any chosen subspace of the configurations (see also Gower, 1995).

Lingoes and Borg (1978) extended GPA to include transformations which do not preserve relative distances between objects. They proposed two models which allow for differential weighting of dimensions. In the first model the dimensions of a "common" space are weighted differentially, while the second model allows for differential weighting of idiosyncratically rotated dimensions of the "common" space. In terms of interpretation, these two models are related to the INDSCAL and IDIOSCAL models of Carroll and Chang (1970, 1972). Lingoes and Borg additionally proposed two so-called vector weighting models where not the dimensions, but the objects of the configurations are allowed to be differentially weighted (see also Borg and Groenen, 1997).

Commandeur (1991) developed convergent algorithms for the estimation of the unknown transformation parameters in the models of Lingoes and Borg. He also generalized these models, including the GPA model, to the situation where the configurations may have different row orders, that is, where information about some objects is missing in some configurations. However, he did not consider the case where configurations have different column orders. Ten Berge, Kiers, and Commandeur (1993) offered an algorithm for GPA of configurations containing missing values in arbitrary places, of which missing rows and columns is a special case.

In the present paper, the matching of M configurations is simultaneously extended in two different directions. First, the general situation is considered where configurations have different column orders. When the match of such configurations is investigated in subspaces of the configurations, the well-known least-squares or Procrustes criterion can be sensitive to "rotating out the variance" into the non-fitted dimensions of the space (see, e.g., Peay, 1988; Dijksterhuis and Gower, 1992; Gower, 1995). Here, a different approach to the least-squares criterion is proposed that does not suffer from this drawback. Also, this criterion preserves relative Euclidean distances in GPA, is computationally efficient, and specializes to the classical least-squares criterion in the situation where a solution is sought in the full space of the configurations. Moreover, when the configurations consist of ratings of objects on variables, the criterion allows for a direct interpretation of the solution in terms of all these variables.

At the same time, in all matching procedures developed below the possibility of performing weighted analyses is incorporated, where the objects are allowed to be weighted with given nonnegative numbers. The analysis of configurations containing different row orders then becomes a special case where object weights corresponding to missing objects are simply set equal to zero. Previous work in this area has been done by Everitt and Gower (1981), and Gower (1995), who proposed a weighted generalized Procrustes analysis method. Groenen, Commandeur and Meulman (1996) discussed a weighted Procrustes analysis involving weighted translation, rotation and dilation of a bootstrap configuration to a given target configuration.

Also, the match between configurations is investigated according to three different models: the GPA model, and the two dimension weighting models proposed by Lingoes and Borg (1978). A discussion of the vector weighting models of Lingoes and Borg is not included, because these two models do not seem very attractive in terms of psychological interpretation and parsimony. Moreover, the corresponding algorithms were found to be prone to convergence to local minima (see Commandeur, 1991). Finally, for all cases a decomposition of the total variance of the data is presented allowing for an evaluation of the separate contributions of objects and configurations to the solution.

In section 3 the weighted GPA of configurations with different column orders is discussed, and in sections 4 and 5 convergent algorithms are developed for the two dimension weighting models of Lingoes and Borg (1978). In section 6 these matching procedures are illustrated with an example, and section 7 gives an overview of the results. First, however, an alternative least-squares criterion for the subspace matching of configurations that will be used throughout the paper is discussed in section 2.

2. An alternative least-squares criterion for the subspace matching of configurations

Ten Berge and Knol (1984) and Peay (1988) discussed how a number of different optimality criteria arise in the matching of M configurations $\mathbf{X}_j$ when different column orders are involved. Letting $\mathbf{R}_j$ denote a columnwise orthonormal matrix ($\mathbf{R}_j'\mathbf{R}_j = \mathbf{I}_{m_j}$) of order $m_j \times m$ where $m \leq \min\{m_j\}$ is the dimensionality of the subspace in which the configurations are matched, define $\mathbf{Z}$ as an $N \times m$ centroid configuration representing the average of the matrices $\mathbf{X}_j\mathbf{R}_j$, that is, $\mathbf{Z} = M^{-1}\sum_j \mathbf{X}_j\mathbf{R}_j$. Then, assuming that all matrices are centered on the origin, Peay (1988) showed that the total variance of the matrices $\mathbf{X}_j$ in a GPA only involving orthonormal transformations can be partitioned as

$$\sum_j \text{tr}\ \mathbf{X}_j'\mathbf{X}_j = M\text{tr}\ \mathbf{Z}'\mathbf{Z} + \sum_j \text{tr}\ (\mathbf{X}_j\mathbf{R}_j - \mathbf{Z})'(\mathbf{X}_j\mathbf{R}_j - \mathbf{Z}) + \sum_j [\text{tr}\ \mathbf{X}_j'\mathbf{X}_j - \text{tr}\ \mathbf{R}_j'\mathbf{X}_j'\mathbf{X}_j\mathbf{R}_j]. \quad (1)$$

The first term on the right hand side of (1) represents the contribution of the centroid configuration to the total variance in m-dimensional subspace, while the second and the third term on the right hand side of (1) are residuals. The second term contains the residual of the solution in m-dimensional subspace, and the third term is the part of the total variance that is lost in the projection in m-dimensional subspace.

In the general situation of different column orders, there are three optimality criteria to consider which potentially all give different solutions for the unknown matrices $\mathbf{R}_j$. The first, the least-squares or Procrustes criterion, consists of the minimization of the first residual term

$$\mathrm{f}(\mathbf{R},\mathbf{Z}) = \sum_j \mathrm{tr}\ (\mathbf{X}_j\mathbf{R}_j - \mathbf{Z})'(\mathbf{X}_j\mathbf{R}_j - \mathbf{Z}) \tag{2}$$

in (1). The second criterion maximizes the term $M\mathrm{tr}\ \mathbf{Z}'\mathbf{Z}$ in (1), and is called the "consensus" criterion by Peay. A third criterion is hidden in the term $M\mathrm{tr}\ \mathbf{Z}'\mathbf{Z}$ in (1) which can be decomposed as follows:

$$M\mathrm{tr}\ \mathbf{Z}'\mathbf{Z} = 2M^{-1}\sum_{i<j} \mathrm{tr}\ \mathbf{R}_i'\mathbf{X}_i'\mathbf{X}_j\mathbf{R}_j + M^{-1}\sum_j \mathrm{tr}\ \mathbf{R}_j'\mathbf{X}_j'\mathbf{X}_j\mathbf{R}_j. \tag{3}$$

In (3) the term $\sum_{i<j} \mathrm{tr}\ \mathbf{R}_i'\mathbf{X}_i'\mathbf{X}_j\mathbf{R}_j$ represents the "inner product" criterion (ten Berge and Knol, 1984). Clearly, if $m_j = m$ for all j (the full-dimensional case), then $\sum_j \mathrm{tr}\ \mathbf{R}_j'\mathbf{X}_j'\mathbf{X}_j\mathbf{R}_j = \sum_j \mathrm{tr}\ \mathbf{X}_j'\mathbf{X}_j$ and all three criteria coincide.

When matching configurations in subspaces least-squares criterion (2) has two important drawbacks. First, as pointed out by Peay (1988), Dijksterhuis and Gower (1992), and Gower (1995), this least-squares criterion has a tendency to minimize the first residual $\sum_j \mathrm{tr}(\mathbf{X}_j\mathbf{R}_j - \mathbf{Z}))'(\mathbf{X}_j\mathbf{R}_j - \mathbf{Z})$ in (1) *by maximizing the second residual* (i.e., $\sum_j[\mathrm{tr}\ \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j - \mathrm{tr}\ \mathbf{R}_j'\mathbf{X}_j'\mathbf{X}_j\mathbf{R}_j]$), that is, it tends to optimize the Procrustes criterion by rotating variance out of m-dimensional subspace. This yields solutions where residual (2) is very small, but where the value of the "consensus" criterion $M\mathrm{tr}\ \mathbf{Z}'\mathbf{Z}$ in (1) is negligible as well (see section 6 for an example).

Second, the literature by now offers four algorithms that can be used to solve (2) over each columnwise orthonormal matrix $\mathbf{R}_j$: the Green-Gower method as discussed in Green and Gower (1979), and ten Berge and Knol (1984), the planar rotation method proposed by Mooijaart and Commandeur (1990), the majorization method of Koshyat and Swayne (1991), and the majorization method proposed by Meulman (1992). However, these methods all have the drawback that they are iterative, and therefore computationally not very efficient.

To ammend these problems while still staying in a least-squares or Procrustes framework, instead of (2) the following alternative least-squares criterion is proposed for the matching of configurations in subspaces of lower dimensionality

$$f(\mathbf{R}) = \sum_{j=1}^{M} \text{tr } (\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j')'(\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j'), \tag{4}$$

where $\mathbf{Z}$ is again defined as $\mathbf{Z} = M^{-1}\sum_j \mathbf{X}_j \mathbf{R}_j$, the average of the matrices $\mathbf{X}_j\mathbf{R}_j$.

The latter formulation essentially differs from the least-squares criterion (2) discussed in ten Berge and Knol (1984) and Peay (1988) in that upward rotations of $\mathbf{Z}$ are performed in (4), while downward rotations of the configurations $\mathbf{X}_j$ are applied in (2). Thus, in (4) a low-dimensional "common" space $\mathbf{Z}$ is determined that, when erected into all higher-dimensional configuration spaces, on the average best matches each and every one of the $\mathbf{X}_j$'s.

Meulman (1998) used such an upward rotation procedure to obtain a distance-based biplot for multidimensional scaling of multivariate data, because this preserves Euclidean distances in the matrix to which the transformation is applied. In (4) also, distances between the objects both in $\mathbf{Z}$ and the $\mathbf{X}_j$'s are preserved. In contrast, the downward rotation method applied to the configurations in (2) distorts the Euclidean distances between the rows of the configurations $\mathbf{X}_j$ *even in a generalized Procrustes analysis* (which supposedly should preserve these distances). Of course, when $m = m_1 = \ldots = m_M$ the two least-squares criteria coincide, as is easily verified.

Another difference between (4) and (2) becomes apparent when the decomposition of the variance in (1) is compared with the one associated with loss function (4):

$$\sum_j \text{tr } \mathbf{X}_j'\mathbf{X}_j = M\text{tr } \mathbf{Z}'\mathbf{Z} + \sum_j \text{tr } (\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j')'(\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j'). \tag{5}$$

In the latter case the problem of rotating out variance can not arise because the minimization of the residual in this decomposition perforce results in the maximization of the "consensus" criterion $M\text{tr } \mathbf{Z}'\mathbf{Z}$. Thus, in (4) the Procrustes and the "consensus" criterion coincide, even when an optimal match is sought in a subspace of the configurations. A third (computational) advantage of using (4) instead of (2) is that each $\mathbf{R}_j$ in (4) can be solved for analytically.

Finally, in the case that the columns of the configurations contain variables, in (4) no additional efforts are required to obtain representations of the variables in the solution. As Meulman (1998) discussed in the context of distance-based multivariate analysis, the matrix products $\mathbf{ZR}_j'$ immediately define a linear biplot with the rows of $\mathbf{Z}$ as points in m-dimensional subspace, and the rows of the matrices $\mathbf{R}_j$ as m-dimensional vectors (connected with the origin) representing the variables of matrices $\mathbf{X}_j$.

In the following sections, least-squares criterion (4) will be used to set up m-dimensional subspace matching procedures for the three above mentioned models. Although the corresponding loss functions are somewhat more involved than (4), the advantages and properties of (4) discussed in the present section go through unchanged.

3. Generalized Procrustes analysis

In generalized Procrustes analysis, the match between configurations is investigated under all relative Euclidean distance preserving transformations. Besides the orthonormal transformations discussed in the previous section, this also includes translations, and isotropic scaling factors. The possiblity of performing object weighted analyses will also be included. Therefore, let $\mathbf{N}_j$ be a diagonal matrix of order $N \times N$ with given nonnegative weights on the diagonal, and $\mathbf{1}$ a $N \times 1$ vector consisting of ones. Also, let s_j be an unknown isotropic scaling factor, $\mathbf{u}_j$ an unknown translation vector of order $m_j \times 1$, and $\mathbf{t}_j$ an unknown translation vector of order $m \times 1$. Let the M unknown translation vectors $\mathbf{u}_j$ be collected in the $\sum_j m_j \times 1$ partitioned vector $\mathbf{u}$, the M unknown translation vectors $\mathbf{t}_j$ in the $Mm \times 1$ partitioned vector $\mathbf{t}$, the isotropic scaling factors in the $M \times 1$ vector $\mathbf{s}$, and the unknown orthonormal matrices in the $\sum_j m_j \times m$ partitioned matrix $\mathbf{R}$, then the alternative least-squares criterion for evaluating the match between M configurations in weighted GPA is defined as:

$$\mathrm{f}(\mathbf{u},\mathbf{t},\mathbf{s},\mathbf{R},\mathbf{Z}) = \sum_{j=1}^{M} \mathrm{tr}\ [s_j(\mathbf{X}_j - \mathbf{1u}_j') - (\mathbf{Z} - \mathbf{1t}_j')\mathbf{R}_j']'\mathbf{N}_j[s_j(\mathbf{X}_j - \mathbf{1u}_j') - (\mathbf{Z} - \mathbf{1t}_j')\mathbf{R}_j']. \qquad (6)$$

In the next subsection it is first discussed how to eliminate the translation vectors **u** and **t** from (6).

Translations in GPA

Defining $\mathbf{A}_j = s_j\mathbf{X}_j - (\mathbf{Z} - \mathbf{1}\mathbf{t}_j')\mathbf{R}_j'$, and only considering one particular translation vector $\mathbf{u}_j$, function (6) can be rewritten as

$$f(\mathbf{u}_j) = s_j^2(\mathbf{1}'\mathbf{N}_j\mathbf{1})\ \mathbf{u}_j'\mathbf{u}_j - 2\ s_j\mathbf{u}_j'\mathbf{A}_j'\mathbf{N}_j\mathbf{1} + d_j = c_j^2\ \mathbf{u}_j'\mathbf{u}_j - 2\ c_j\mathbf{u}_j'\mathbf{y}_j + d_j, \tag{7}$$

where d_j is a term independent of $\mathbf{u}_j$, $\mathbf{y}_j = \mathbf{A}_j'\mathbf{N}_j\mathbf{1}/(\mathbf{1}'\mathbf{N}_j\mathbf{1})^{1/2}$, and $c_j = s_j(\mathbf{1}'\mathbf{N}_j\mathbf{1})^{1/2}$. Again rewriting the latter function, it satisfies

$$f(\mathbf{u}_j) = \|c_j\mathbf{u}_j - \mathbf{y}_j\|^2 + d_j - \mathbf{y}_j'\mathbf{y}_j \geq d_j - \mathbf{y}_j'\mathbf{y}_j.$$

This lower bound, and hence the global minimum, is attained for $c_j\mathbf{u}_j - \mathbf{y}_j = \mathbf{0}$, and thus for

$$\mathbf{u}_j = c_j^{-1}\mathbf{y}_j = \frac{\mathbf{A}_j'\mathbf{N}_j\mathbf{1}}{s_j(\mathbf{1}'\mathbf{N}_j\mathbf{1})}. \tag{8}$$

Substitution of (8) in (6) gives

$$f(\mathbf{s},\mathbf{R},\mathbf{Z}) = \sum_{j=1}^{M} \mathrm{tr}\ (s_j\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j')'\mathbf{C}_j(s_j\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j'), \tag{9}$$

where $\mathbf{C}_j = \mathbf{N}_j\mathbf{V}_j$ with $\mathbf{V}_j = (\mathbf{I} - \frac{\mathbf{1}\mathbf{1}'\mathbf{N}_j}{\mathbf{1}'\mathbf{N}_j\mathbf{1}})$. Result (9) implies that the translation problem in GPA is taken care of by weight-centering matrices $\mathbf{X}_j$ and $\mathbf{Z}\mathbf{R}_j'$ on the origin of m_j-dimensional space. It may be noted that $\mathbf{C}_j = \mathbf{N}_j\mathbf{V}_j = \mathbf{V}_j'\mathbf{N}_j\mathbf{V}_j$ is a symmetric matrix, as is readily verified, and that $\mathbf{C}_j\mathbf{1} = \mathbf{0}$, meaning that $\mathbf{C}_j$ is always a singular matrix.

In the next sections it is first discussed how configuration **Z** can be eliminated from (6), since the resulting loss function allows for an efficient update strategy for the isotropic scaling factors, and for square orthonormal matrices $\mathbf{R}_j$. Then it is shown how to obtain optimal (square, and columnwise) orthonormal matrices $\mathbf{R}_j$, and nontrivial isotropic scaling factors s_j.

The centroid configuration in GPA

Here, we discuss how to minimize (9) with respect to $\mathbf{Z}$ after the translation vectors have been eliminated from (6), and then show that $\mathbf{Z}$ itself can also be eliminated from the loss function. Defining $\mathbf{A}_j = s_j\mathbf{X}_j$, loss function (9) may be written as

$$\mathrm{f}(\mathbf{Z}) = \mathrm{c} + \sum_{j=1}^{M} \mathrm{tr}\, \mathbf{Z}'\mathbf{C}_j\mathbf{Z} - 2 \sum_{j=1}^{M} \mathrm{tr}\, \mathbf{Z}'\mathbf{C}_j\mathbf{A}_j\mathbf{R}_j = \mathrm{d} + \sum_{j=1}^{M} \|\mathbf{F}_j\mathbf{Z} - \mathbf{F}_j\mathbf{A}_j\mathbf{R}_j\|^2 \tag{10}$$

with $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$, and where the terms $\mathrm{c} = \sum_j \mathrm{tr}\, \mathbf{A}_j'\mathbf{C}_j\mathbf{A}_j$ and $\mathrm{d} = \mathrm{c} - \sum_j \mathrm{tr}\, \mathbf{R}_j'\mathbf{A}_j'\mathbf{C}_j\mathbf{A}_j\mathbf{R}_j$ are both independent of $\mathbf{Z}$. Let the matrices $\mathbf{F}_j$ be collected in the partitioned matrix $\mathbf{F}$ of order $MN{\times}N$, and let the matrices $\mathbf{B}_j = \mathbf{F}_j\mathbf{A}_j\mathbf{R}_j$ be collected in the partitioned matrix $\mathbf{B}$ of order $MN{\times}m$. Then loss function (10) may be rewritten as

$$\mathrm{f}(\mathbf{Z}) = \mathrm{d} + \|\mathbf{FZ} - \mathbf{B}\|^2. \tag{11}$$

To minimize (11) with respect to $\mathbf{Z}$ is the classical multivariate multiple regression problem. The general solution of this problem is well-known, and follows from the normal equations corresponding to (11):

$$\mathbf{F}'\mathbf{F}\mathbf{Z} = \mathbf{F}'\mathbf{B}. \tag{12}$$

Re-expressing (12) in the original matrices and scalars yields

$$\mathbf{CZ} = \sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j, \tag{13}$$

with $\mathbf{C} = \sum_j\mathbf{C}_j$. Hence, the global minimum of (9) with respect to $\mathbf{Z}$ is attained for

$$\mathbf{Z} = \mathbf{C}^{-}\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j, \tag{14}$$

where $\mathbf{C}^{-}$ is the Moore-Penrose inverse of the sum of the matrices $\mathbf{C}_j$.

The reason that a generalized inverse must be determined in order to solve (13) for unknown $\mathbf{Z}$ is that

$$\mathbf{C1} = \sum_{j=1}^{M} \mathbf{N}_j\mathbf{V}_j\mathbf{1} = \sum_{j=1}^{M} \mathbf{N}_j\mathbf{0} = \mathbf{0},$$

meaning that the sum of the $\mathbf{C}_j$'s has no proper inverse.

Substitution of (14) into (9) yields a loss function that no longer contains the centroid configuration:

$$\mathrm{f}(\mathbf{s},\mathbf{R}) = \sum_{j=1}^{M} s_j^2 \operatorname{tr} \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j - \operatorname{tr} (\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j)'\mathbf{C}^{-}(\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j), \qquad (15)$$

leaving only two sets of unknowns. As will be shown below, formulation (15) of the weighted GPA loss function is particularly efficient for obtaining updates for the isotropic scaling factors **s**, as well as to determine updates for orthonormal matrices $\mathbf{R}_j$ that are square. However, if $\mathbf{R}_j$ is a rectangular, columnwise orthonormal matrix, then formulation (9) yields much more efficient results.

Orthonormal transformations

For fixed isotropic scaling factors **s**, the minimization of (15) is equivalent to the maximization of

$$\mathrm{g}(\mathbf{R}) = \operatorname{tr} (\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j)'\mathbf{C}^{-}(\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j) \qquad (16)$$

subject to $\mathbf{R}_j'\mathbf{R}_j = \mathbf{I}_m$ for $j = 1, \ldots, M$. Letting $\mathbf{A}_j = s_j\mathbf{X}_j$ and considering only one particular $\mathbf{R}_j$, (16) can be written as

$$\mathrm{f}(\mathbf{R}_j) = \operatorname{tr} \mathbf{R}_j'\mathbf{D}_j\mathbf{R}_j + 2 \operatorname{tr} \mathbf{R}_j'\mathbf{B}_j + e_j, \qquad (17)$$

where $\mathbf{D}_j = \mathbf{A}_j'\mathbf{C}_j\mathbf{C}^{-}\mathbf{C}_j\mathbf{A}_j$, $\mathbf{B}_j = \mathbf{A}_j'\mathbf{C}_j\mathbf{C}^{-}(\sum_{i \neq j} \mathbf{C}_i\mathbf{A}_i\mathbf{R}_i)$, and e_j is a term independent of $\mathbf{R}_j$. When $\mathbf{R}_j$ is of order $m_j \times m$ with $m < m_j$, which is the general situation, the maximization of (17) requires an iterative procedure, which is computationally not very efficient (see section 2).

However, in the special situation that $m_j = m$ the maximization of (17) has the following well-known analytical solution. The maximization of (17) is then equivalent to the maximization of

$$h(\mathbf{R}_j) = \text{tr } \mathbf{R}_j'\mathbf{B}_j \qquad (18)$$

subject to $\mathbf{R}_j'\mathbf{R}_j = \mathbf{I}_m$. If $\mathbf{B}_j = \mathbf{P}_j\mathbf{\Phi}_j\mathbf{Q}_j'$ is a singular value decomposition of $\mathbf{B}_j$, then $\mathbf{R}_j = \mathbf{P}_j\mathbf{Q}_j'$ maximizes (18) subject to $\mathbf{R}_j'\mathbf{R}_j = \mathbf{I}_m$ (see, Cliff, 1966; Schönemann, 1966).

In the general situation where $m < m_j$, iterative procedures for the minimization of (17) can be avoided by switching back to formulation (9) of the optimization problem at hand. Since, for fixed $\mathbf{s}$ and $\mathbf{Z}$, and considering only one $\mathbf{R}_j$, (9) can be written as

$$f(\mathbf{R}_j) = c - 2 \text{ tr } \mathbf{R}_j'\mathbf{A}_j'\mathbf{C}_j\mathbf{Z} = c - 2 \text{ tr } \mathbf{R}_j'\mathbf{B}_j, \qquad (19)$$

with c a term independent of $\mathbf{R}_j$ and $\mathbf{B}_j = \mathbf{A}_j'\mathbf{C}_j\mathbf{Z}$, the efficient analytical solution for (18) can be applied to (19) also.

Isotropic scaling factors

To determine optimal isotropic scaling factors keeping $\mathbf{R}$ fixed, some kind of restriction must be imposed upon the isotropic scaling factors because the ensuing algorithm will otherwise converge to the trivial solution $\mathbf{s} = \mathbf{0}$. The isotropic scaling factors are restricted to satisfy

$$\sum_{j=1}^{M} s_j^2 \text{ tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j = \sum_{j=1}^{M} \text{tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j. \qquad (20)$$

This side condition is a generalization of the constraint that Gower (1975) used in the unweighted case with equal column orders. It may be noted that the least-squares criterion proposed in the present paper avoids an ambiguity that arises in the matching in subspaces of the configurations using Procrustes criterion (2). In the latter case, a choice has to be made whether to apply constraint (20) to the isotropic scaling factors, or the constraint $\sum_j s_j^2$ tr $\mathbf{R}_j'\mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j = \sum_j \text{tr } \mathbf{R}_j'\mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j$ (see Gower, 1995).

Let $\mathbf{A}_j = \mathbf{X}_j\mathbf{R}_j$, and matrix $\mathbf{Y}$ of order $M{\times}M$ be defined as

$$\mathbf{Y} = \begin{bmatrix} \text{tr } \mathbf{A}_1'\mathbf{C}_1\mathbf{C}^-\mathbf{C}_1\mathbf{A}_1 & \text{tr } \mathbf{A}_1'\mathbf{C}_1\mathbf{C}^-\mathbf{C}_2\mathbf{A}_2 & \dots & \text{tr } \mathbf{A}_1'\mathbf{C}_1\mathbf{C}^-\mathbf{C}_M\mathbf{A}_M \\ \text{tr } \mathbf{A}_2'\mathbf{C}_2\mathbf{C}^-\mathbf{C}_1\mathbf{A}_1 & \text{tr } \mathbf{A}_2'\mathbf{C}_2\mathbf{C}^-\mathbf{C}_2\mathbf{A}_2 & \dots & \text{tr } \mathbf{A}_2'\mathbf{C}_2\mathbf{C}^-\mathbf{C}_M\mathbf{A}_M \\ \vdots & \vdots & & \vdots \\ \text{tr } \mathbf{A}_M'\mathbf{C}_M\mathbf{C}^-\mathbf{C}_1\mathbf{A}_1 & \text{tr } \mathbf{A}_M'\mathbf{C}_M\mathbf{C}^-\mathbf{C}_2\mathbf{A}_2 & \dots & \text{tr } \mathbf{A}_M'\mathbf{C}_M\mathbf{C}^-\mathbf{C}_M\mathbf{A}_M \end{bmatrix}. \tag{21}$$

Also, define matrix $\mathbf{N}$ of order $M{\times}M$ as

$$\mathbf{N} = \begin{bmatrix} \text{tr } \mathbf{X}_1'\mathbf{C}_1\mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \text{tr } \mathbf{X}_2'\mathbf{C}_2\mathbf{X}_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \text{tr } \mathbf{X}_M'\mathbf{C}_M\mathbf{X}_M \end{bmatrix}. \tag{22}$$

Then (15) can be written as

$$\text{f}(\mathbf{s}) = \sum_{j=1}^{M} s_j^2 \text{ tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j - \text{tr } (\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{A}_j)'\mathbf{C}^-(\sum_{j=1}^{M} s_j\mathbf{C}_j\mathbf{A}_j)$$

$$= \mathbf{1}'\mathbf{N}\mathbf{1} - \mathbf{s}'\mathbf{Y}\mathbf{s}, \tag{23}$$

on the condition that the scaling factors satisfy (20), which can be written as

$$\mathbf{s}'\mathbf{N}\mathbf{s} = \mathbf{1}'\mathbf{N}\mathbf{1}. \tag{24}$$

Therefore, the minimization of (23) is equivalent to the maximization of

$$\text{h}(\mathbf{s}) = \mathbf{s}'\mathbf{Y}\mathbf{s}. \tag{25}$$

subject to (24). Let

$$\mathbf{N}^{-1/2}\mathbf{Y}\mathbf{N}^{-1/2} = \mathbf{P}\boldsymbol{\Phi}\mathbf{P}' \tag{26}$$

be an eigenvector-eigenvalue decomposition of $\mathbf{N}^{-1/2}\mathbf{Y}\mathbf{N}^{-1/2}$, with eigenvalues ordered in non-increasing order on the diagonal of $\boldsymbol{\Phi}$, and define $\mathbf{p}_1$ as the first column of $\mathbf{P}$. Then the isotropic scaling factors minimizing (15) subject to (20) are

$$\mathbf{s} = (\mathbf{1}'\mathbf{N}\mathbf{1})^{1/2}\mathbf{N}^{-1/2}\mathbf{p}_1, \tag{27}$$

where $\mathbf{N}$ and $\mathbf{p}_1$ are defined in (22) and (26), respectively (for a proof, see ten Berge, 1977; ten Berge and Bekker, 1993).

Concluding, for fixed $\mathbf{R}$ an analytical solution is available for the unknown isotropic scaling constants s which is guaranteed to yield the global optimum of the GPA loss function and does not involve $\mathbf{Z}$ in any way. For fixed s, the best strategy for obtaining efficient updates for orthonormal matrices $\mathbf{R}_j$ depends on their form: If they are square orthonormal then formulation (16) gives the most efficient results, if they are columnwise orthonormal then formulation (19) should be used.

Thus, to minimize the weighted GPA loss function (15) with respect to $\mathbf{R}$ and s, or (9) with respect to $\mathbf{R}$, s, and $\mathbf{Z}$, an iterative procedure must be used. Since (9) and (15) are bounded below by zero, this algorithm must converge, although not necessarily to the global minimum.

Non-uniqueness of the GPA solution and analysis of variance

The GPA solution is only unique up to a simultaneous rotation of the centroid configuration $\mathbf{Z}$, because the value of the GPA loss function is unaffected by using:

$$f(\mathbf{R},\mathrm{s},\mathbf{Z}) = \sum_{j=1}^{M} \mathrm{tr}\ (\mathrm{s}_j\mathbf{X}_j - \bar{\mathbf{Z}}\bar{\mathbf{R}}_j')'\mathbf{C}_j(\mathrm{s}_j\mathbf{X}_j - \bar{\mathbf{Z}}\bar{\mathbf{R}}_j'), \tag{28}$$

where $\bar{\mathbf{Z}} = \mathbf{ZK}$, $\bar{\mathbf{R}}_j = \mathbf{R}_j\mathbf{K}$, and $\mathbf{K}$ is an arbitrary orthonormal matrix of order $m \times m$. As is usual in GPA, at convergence this indeterminacy is exploited to rotate the whole solution to the principal components of $\mathbf{Z}$. Thus, the eigenvalue-eigenvector decomposition $\mathbf{Z}'\mathbf{CZ} = \mathbf{K\Lambda K}'$ is determined, and the centroid configuration is set equal to $\mathbf{ZK}$, and the orthonormal matrices $\bar{\mathbf{R}}_j$ are set equal to $\bar{\mathbf{R}}_j = \mathbf{R}_j\mathbf{K}$ for each j.

Generalizing (5) to weighted GPA the following decomposition of the total variance is obtained

$$\textstyle\sum_j \mathrm{tr}\ \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j = \mathrm{tr}\ \mathbf{Z}'\mathbf{CZ} + \sum_j \mathrm{tr}\ (\mathrm{s}_j\mathbf{X}_j - \mathbf{ZR}_j')'\mathbf{C}_j(\mathrm{s}_j\mathbf{X}_j - \mathbf{ZR}_j'). \tag{29}$$

The residual term in (29) can be partitioned such that the contribution of each separate object and each separate configuration to the total residual sum of squares can be assessed. Let $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$, and $\mathbf{A}_j = \mathbf{F}_j(s_j\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j')(s_j\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j')'\mathbf{F}_j'$. Then the vector $(\text{diag}\mathbf{A}_j)\mathbf{1}$ contains the contribution of each object i in configuration j to the residual term, while the vector $\sum_j(\text{diag}\mathbf{A}_j)\mathbf{1}$ consists of the residual contribution of each object over all configurations simultaneously, and the scalar $\mathbf{1}'(\text{diag}\mathbf{A}_j)\mathbf{1}$ contains the residual contribution of each configuration j. The m eigenvalues $\mathbf{\Lambda}$ of the eigenvalue decomposition $\mathbf{Z}'\mathbf{C}\mathbf{Z} = \mathbf{K}\mathbf{\Lambda}\mathbf{K}'$ indicate the contributions of each dimension of $\mathbf{ZK}$ to the GPA solution.

4. Dimension weighting

In contrast with the GPA model, where only uniform rescaling of the configurations is considered, in dimension weighting the m dimensions of an unknown $N\times m$ centroid configuration $\mathbf{Y}$ are allowed to be differentially weighted, with $m \leq \min\{m_j\}$. To determine the unknown parameters in the first dimension weighting model of Lingoes and Borg (1978) the following least-squares loss function is defined:

$$\text{f}(\mathbf{g},\mathbf{h},\mathbf{Q},\mathbf{W},\mathbf{Y}) =$$

$$= \sum_{j=1}^{M} \text{tr}\ [(\mathbf{X}_j - \mathbf{1}\mathbf{g}_j') - (\mathbf{Y} - \mathbf{1}\mathbf{h}_j')\mathbf{W}_j\mathbf{Q}_j']'\mathbf{N}_j[(\mathbf{X}_j - \mathbf{1}\mathbf{g}_j') - (\mathbf{Y} - \mathbf{1}\mathbf{h}_j')\mathbf{W}_j\mathbf{Q}_j']. \tag{30}$$

In (30) the unknown translation vectors $\mathbf{g}_j$ and $\mathbf{h}_j$ of order $m_j\times 1$ and $m\times 1$ are assumed to be collected in the $\sum_j m_j\times 1$ and $Mm\times 1$ vectors $\mathbf{g}$ and $\mathbf{h}$, respectively, the unknown $m_j\times m$ columnwise orthonormal matrices $\mathbf{Q}_j$ in the $\sum_j m_j\times m$ partitioned matrix $\mathbf{Q}$, and the unknown diagonal dimension weight matrices $\mathbf{W}_j$ in the $Mm\times m$ partitioned matrix $\mathbf{W}$ (see also Commandeur, 1991).

First, the translation vectors $\mathbf{g}_j$ optimizing (30) are determined, since this makes it possible to simultaneously eliminate $\mathbf{g}$ and $\mathbf{h}$ from (30). Next, it is shown how to minimize the simplified loss function with respect to the $m_j\times m$ orthonormal matrices $\mathbf{Q}_j$, the $m\times m$ dimension weight matrices $\mathbf{W}_j$ and the centroid configuration $\mathbf{Y}$, respectively. Then the non-uniqueness properties of this model are discussed, as well as an analysis of variance making it possible to evaluate the separate contributions of configurations and objects to the solution.

Translations

Defining $\mathbf{A}_j = \mathbf{X}_j - (\mathbf{Y} - \mathbf{1h}_j')\mathbf{W}_j\mathbf{Q}_j'$, and considering only one particular $\mathbf{g}_j$, (30) may be written as

$$f(\mathbf{g}_j) = \mathbf{1}'\mathbf{N}_j\mathbf{1}\,\mathbf{g}_j'\mathbf{g}_j - 2\,\mathbf{g}_j'\mathbf{A}_j'\mathbf{N}_j\mathbf{1} + d_j$$

$$= c_j^2\,\mathbf{g}_j'\mathbf{g}_j - 2\,c_j\mathbf{g}_j'\mathbf{b}_j + d_j, \tag{31}$$

where d_j is a term independent of $\mathbf{g}_j$, $\mathbf{b}_j \equiv \mathbf{A}_j'\mathbf{N}_j\mathbf{1}/(\mathbf{1}'\mathbf{N}_j\mathbf{1})^{1/2}$, and $c_j \equiv (\mathbf{1}'\mathbf{N}_j\mathbf{1})^{1/2}$. Applying the same procedure to (31) as discussed for GPA, the global minimum of (31) is attained for

$$\mathbf{g}_j = \frac{\mathbf{A}_j'\mathbf{N}_j\mathbf{1}}{\mathbf{1}'\mathbf{N}_j\mathbf{1}}. \tag{32}$$

Substitution of (32) in (30) yields

$$f(\mathbf{Q},\mathbf{W},\mathbf{Y}) = \sum_{j=1}^{M} \mathrm{tr}\,(\mathbf{X}_j - \mathbf{Y}\mathbf{W}_j\mathbf{Q}_j')'\mathbf{C}_j(\mathbf{X}_j - \mathbf{Y}\mathbf{W}_j\mathbf{Q}_j'). \tag{33}$$

where again $\mathbf{C}_j = \mathbf{V}_j'\mathbf{N}_j\mathbf{V}_j = \mathbf{N}_j\mathbf{V}_j$, with $\mathbf{V}_j = (\mathbf{I} - \frac{\mathbf{11}'\mathbf{N}_j}{\mathbf{1}'\mathbf{N}_j\mathbf{1}})$. Substitution of (32) in (30) has the effect of simultaneously eliminating $\mathbf{g}$ and $\mathbf{h}$ from (30), showing that, just like in GPA, the problem of translations in the dimension weighting model is implicitly solved by weight-centering matrices $\mathbf{X}_j$ and $\mathbf{Y}$ on the origin of m-dimensional space.

Orthonormal transformations

To determine the orthonormal matrices $\mathbf{Q}_j$ minimizing (33), define $\mathbf{A}_j = \mathbf{YW}_j$ and rewrite (33) as

$$f(\mathbf{Q}) = \sum_{j=1}^{M} \mathrm{tr}\,(\mathbf{X}_j - \mathbf{A}_j\mathbf{Q}_j')'\mathbf{C}_j(\mathbf{X}_j - \mathbf{A}_j\mathbf{Q}_j'). \tag{34}$$

Considering only one particular $\mathbf{Q}_j$, (34) can be written as

$$f(\mathbf{Q}_j) = d_j - 2 \text{ tr } \mathbf{Q}_j'\mathbf{X}_j'\mathbf{C}_j\mathbf{A}_j, \tag{35}$$

where d_j is a term independent of $\mathbf{Q}_j$. This problem is identical to the one encountered in GPA, and has an analytical solution.

Dimension weights

For fixed $\mathbf{Q}$ and $\mathbf{Y}$, the problem is how to minimize

$$f(\mathbf{W}) = \sum_{j=1}^{M} \text{tr } (\mathbf{X}_j - \mathbf{Y}\mathbf{W}_j\mathbf{Q}_j')'\mathbf{C}_j(\mathbf{X}_j - \mathbf{Y}\mathbf{W}_j\mathbf{Q}_j')$$

$$= c + \sum_{j=1}^{M} [\text{tr } \mathbf{W}_j^2\mathbf{Y}'\mathbf{C}_j\mathbf{Y} - 2 \text{ tr } \mathbf{W}_j\mathbf{Y}'\mathbf{C}_j\mathbf{X}_j\mathbf{Q}_j] = c + \sum_{j=1}^{M} [d_j + \|\mathbf{W}_j\mathbf{A}_j^{1/2} - \mathbf{A}_j^{-1/2}\mathbf{B}_j\|^2], \tag{36}$$

over diagonal matrices $\mathbf{W}_j$, where $\mathbf{A}_j = (\text{diag } \mathbf{Y}'\mathbf{C}_j\mathbf{Y})$, $\mathbf{B}_j = (\text{diag } \mathbf{Y}'\mathbf{C}_j\mathbf{X}_j\mathbf{Q}_j)$, and c and d_j are both terms independent of $\mathbf{W}$. The global minimum of (36) is attained by computing

$$\mathbf{W}_j = \mathbf{A}_j^{-1}\mathbf{B}_j = (\text{diag } \mathbf{Y}'\mathbf{C}_j\mathbf{Y})^{-1}(\text{diag } \mathbf{Y}'\mathbf{C}_j\mathbf{X}_j\mathbf{Q}_j) \tag{37}$$

for $j = 1, \ldots, m$.

The centroid configuration

For fixed $\mathbf{Q}$ and $\mathbf{W}$, the minimum of

$$f(\mathbf{Y}) = \Sigma_j \text{ tr } (\mathbf{A}_j - \mathbf{F}_j\mathbf{Y}\mathbf{B}_j)'(\mathbf{A}_j - \mathbf{F}_j\mathbf{Y}\mathbf{B}_j) \tag{38}$$

needs to be determined, where $\mathbf{A}_j = \mathbf{F}_j\mathbf{X}_j$, $\mathbf{B}_j = \mathbf{W}_j\mathbf{Q}_j'$, and $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$. Define $\mathbf{a}_j = (\text{vec } \mathbf{A}_j)$, that is, let $\mathbf{a}_j$ be the $Nm_j\times 1$ vector consisting of the columns of $\mathbf{A}_j$ stacked on top of one another, and let (vec $\mathbf{F}_j\mathbf{Y}\mathbf{B}_j$) denote the $Nm_j\times 1$ vector which is obtained by stacking the columns of $\mathbf{F}_j\mathbf{Y}\mathbf{B}_j$ on top of one another. Since

$$(\text{vec } \mathbf{F}_j\mathbf{Y}\mathbf{B}_j) = (\mathbf{B}_j' \otimes \mathbf{F}_j)(\text{vec } \mathbf{Y}),$$

where (vec **Y**) is the $Nm \times 1$ vector containing the columns of **Y** stacked on top of one another and $\otimes$ denotes the right Kronecker product, (38) may written as

$$f(\mathbf{y}) = \sum_{j=1}^{M} (\mathbf{a}_j - \mathbf{D}_j\mathbf{y})'(\mathbf{a}_j - \mathbf{D}_j\mathbf{y}), \tag{39}$$

where $\mathbf{y}$ = (vec **Y**) of order $Nm \times 1$, and $\mathbf{D}_j = (\mathbf{B}_j' \otimes \mathbf{F}_j)$ of order $Nm_j \times Nm$. Letting the partitioned vector **a** of order $MNm_j \times 1$ contain the $\mathbf{a}_j$'s stacked on top of one another, and collecting the matrices $\mathbf{D}_j$ in the supermatrix **D** of order $MNm_j \times Nm$, (39) may be written as

$$f(\mathbf{y}) = (\mathbf{a} - \mathbf{D}\mathbf{y})'(\mathbf{a} - \mathbf{D}\mathbf{y}). \tag{40}$$

This is the classical univariate multiple regression problem. The solution follows from the normal equations

$$\mathbf{D}'\mathbf{D}\mathbf{y} = \mathbf{D}'\mathbf{a},$$

which may be re-expressed in terms of the original matrices as

$$[\sum_{j=1}^{M} (\mathbf{W}_j^2 \otimes \mathbf{C}_j)](\text{vec } \mathbf{Y}) = \sum_{j=1}^{M} (\text{vec } \mathbf{C}_j\mathbf{X}_j\mathbf{Q}_j\mathbf{W}_j). \tag{41}$$

Hence, the global minimum of (38) is attained for

$$\text{vec } \mathbf{Y} = [\sum_{j=1}^{M} (\mathbf{W}_j^2 \otimes \mathbf{C}_j)]^{-}[\sum_{j=1}^{M} (\text{vec } \mathbf{C}_j\mathbf{X}_j\mathbf{Q}_j\mathbf{W}_j)], \tag{42}$$

where $[\Sigma_j (\mathbf{W}_j^2 \otimes \mathbf{C}_j)]^-$ is the Moore-Penrose generalized inverse of $\Sigma_j (\mathbf{W}_j^2 \otimes \mathbf{C}_j)$. The latter matrix of order $Nm \times Nm$ is block-diagonal, and the Moore-Penrose generalized inverse of this matrix can therefore be computed for each of the m symmetric blocks of order $N \times N$ separately.

In the special case that all object weights are equal to one, and assuming that the configurations have been centered on the origin, (42) simplifies into

$$\mathbf{Y} = (\sum_{j=1}^{M} \mathbf{X}_j\mathbf{Q}_j\mathbf{W}_j)(\sum_{j=1}^{M} \mathbf{W}_j^2)^{-1}, \tag{43}$$

which only requires the simple computation of the inverse of an $m \times m$ diagonal matrix.

Summarizing, a convergent algorithm for the minimization of (33) is obtained by alternatingly computing new $\mathbf{Q}_j$'s minimizing (35), new updates for the dimension weights $\mathbf{W}_j$ using (37), and a new update for the centroid configuration $\mathbf{Y}$ according to (42). The algorithm may be started with $\mathbf{Y} = \mathbf{Z}$, where $\mathbf{Z}$ is the optimal centroid configuration obtained with GPA, and $\mathbf{Q} = \mathbf{R}$ where $\mathbf{R}$ is the partitioned matrix containing the optimal orthonormal matrices $\mathbf{R}_j$ obtained with GPA.

Non-uniqueness of the solution and analysis of variance

The solution for the dimension weighting model is only unique up to a weighting of the optimal dimensions weights $\mathbf{W}_j$ together with an inverse weighting of the columns of the optimal centroid $\mathbf{Y}$. This is true because all the following solutions are equivalent for fixed $\mathbf{Q}$, $\mathbf{W}$, and $\mathbf{Y}$:

$$\mathrm{f}(\mathbf{Q},\mathbf{W},\mathbf{Y}) = \sum_{j=1}^{M} \mathrm{tr}\,(\mathbf{X}_j - \bar{\mathbf{Y}}\bar{\mathbf{W}}_j\mathbf{Q}_j')'\mathbf{C}_j(\mathbf{X}_j - \bar{\mathbf{Y}}\bar{\mathbf{W}}_j\mathbf{Q}_j'), \tag{44}$$

where $\bar{\mathbf{Y}} = \mathbf{Y}\mathbf{L}^{-1}$, $\bar{\mathbf{W}}_j = \mathbf{L}\mathbf{W}_j$, and $\mathbf{L}$ is an arbitrary diagonal matrix of order $m \times m$.

This indeterminacy can be used to choose the following diagonal matrix $\mathbf{L}$ with attractive properties:

$$\mathbf{L} = (\mathrm{diag}\ \mathbf{Y}'\mathbf{Y})^{1/2}, \tag{45}$$

yielding a weighted solution where

$$\bar{\mathbf{Y}} = \mathbf{Y}\mathbf{L}^{-1} = \mathbf{Y}(\mathrm{diag}\ \mathbf{Y}'\mathbf{Y})^{-1/2} \tag{46}$$

and

$$\bar{\mathbf{W}}_j = \mathbf{L}\mathbf{W}_j = (\mathrm{diag}\ \mathbf{Y}'\mathbf{Y})^{1/2}\mathbf{W}_j \tag{47}$$

for $j = 1, \ldots, M$.

Postmultiplying $\mathbf{Y}$ with $\mathbf{L}^{-1}$ has the effect of unit normalizing the columns of $\mathbf{Y}$ because $(\mathrm{diag}\ \mathbf{L}^{-1}\mathbf{Y}'\mathbf{Y}\mathbf{L}^{-1}) = \mathbf{I}_m$. Moreover, in the unweighted case (where $\mathbf{C}_j$ is the usual centering matrix $\mathbf{J} = (\mathbf{I} - \mathbf{1}\mathbf{1}'/\mathbf{1}'\mathbf{1})$ for $j = 1, \ldots, M$) it can be proven that the choice for (45) guarantees that $(\mathbf{1}'\bar{\mathbf{W}}_j^2\mathbf{1})/(\text{tr }\mathbf{X}_j'\mathbf{X}_j)$ is equal to the squared correlation between the elements of $\mathbf{X}_j$ and the elements of $\mathbf{YW}_j\mathbf{Q}_j'$ (see Commandeur, 1991). Also, both in the weighted and unweighted case the contribution of a configuration to the solution can be assessed with

$$r^2(\mathbf{F}_j\mathbf{X}_j,\mathbf{F}_j\mathbf{YW}_j\mathbf{Q}_j') = \frac{\text{tr }\mathbf{W}_j\mathbf{Y}'\mathbf{C}_j\mathbf{YW}_j}{\text{tr }\mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j}, \tag{48}$$

where $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$.

A property of the dimension weighting solution is that the centroid configuration $\mathbf{Y}$ has optimally rotated dimensions since these are the ones that are differentially weighted. Analogous to the INDSCAL model, therefore, in this case also it is the model itself that dictates, as it were, which axes are to be used for interpretation.

The total variance in the configurations can be partitioned as follows. Applying decomposition (5) to the present model yields:

$$\textstyle\sum_j \text{tr }\mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j = \sum_j \text{tr }\mathbf{W}_j\mathbf{Y}'\mathbf{C}_j\mathbf{YW}_j + \sum_j \text{tr }(\mathbf{X}_j - \mathbf{YW}_j\mathbf{Q}_j')'\mathbf{C}_j(\mathbf{X}_j - \mathbf{YW}_j\mathbf{Q}_j'). \tag{49}$$

The first term on the right-hand side of (49) represents the contribution of the weighted centroid configuration to the total variance in m-dimensional subspace, while the second term on the right hand side of (49) gives the residual variance. Additivity of the first and second terms on the right-hand side of (49) is guaranteed because of the orthogonality of the predictor and residual space in regression which applies to the determination of an optimal $\mathbf{Y}$ in (40). Again, since the minimization of the residual variance in (49) amounts to the maximization of the consensus criterion $\sum_j \text{tr }\mathbf{W}_j\mathbf{Y}'\mathbf{C}_j\mathbf{YW}_j$, there is no danger of rotating out variance into the non-fitted dimensions of the space.

Analogous to the GPA model, the residual term in (49) can be further partitioned such that the contribution of each separate object and each separate configuration to the overall residual variance can be assessed. Let $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$, and $\mathbf{A}_j = \mathbf{F}_j(\mathbf{X}_j - \mathbf{YW}_j\mathbf{Q}_j')(\mathbf{X}_j - \mathbf{YW}_j\mathbf{Q}_j')'\mathbf{F}_j'$. Then the vector $(\text{diag}\mathbf{A}_j)\mathbf{1}$ contains the contribution of each object in configuration j to the residual variance, the vector $\sum_j(\text{diag}\mathbf{A}_j)\mathbf{1}$ has the residual contributions of the objects over all configurations simultaneously, and the scalar $\mathbf{1}'(\text{diag}\mathbf{A}_j)\mathbf{1}$ contains the residual contribution of each configuration j in m-dimensional (sub)space. The diagonal elements of $\sum_j \mathbf{W}_j \mathbf{Y}'\mathbf{C}_j \mathbf{Y}\mathbf{W}_j$ can be used to assess the contribution of each dimension to the solution.

5. Dimension weighting with idiosyncratic rotations

To determine the optimal translations, orthonormal transformations, dimension weights, and centroid configuration in the second dimension weighting model with idiosyncratic rotations of Lingoes and Borg (1978), the following least-squares loss function is defined:

$$f(\mathbf{g},\mathbf{h},\mathbf{Q},\mathbf{S},\mathbf{W},\mathbf{Y}) =$$

$$\sum_{j=1}^{M} [(\mathbf{X}_j - \mathbf{1}\mathbf{g}_j') - (\mathbf{Y} - \mathbf{1}\mathbf{h}_j')\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j']'\mathbf{N}_j[(\mathbf{X}_j - \mathbf{1}\mathbf{g}_j') - (\mathbf{Y} - \mathbf{1}\mathbf{h}_j')\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j']. \tag{50}$$

In (50), vectors $\mathbf{g}$ and $\mathbf{h}$ and matrices $\mathbf{Q}$ and $\mathbf{W}$ are defined as before, while $\mathbf{S}$ is a partitioned matrix of order $Mm \times m$ in which the M orthonormal matrices $\mathbf{S}_j$ are collected. Matrices $\mathbf{X}_j$ and $\mathbf{N}_j$ are given, and also defined as before. The essential difference between loss functions (50) and (30) consists of the rotation matrices $\mathbf{S}_j$, allowing for the weighting of idiosyncratically rotated dimensions of $\mathbf{Y}$.

The treatment of the translation vectors $\mathbf{g}_j$ and $\mathbf{h}_j$ in (50) is so similar to the one described for the previous dimension weighting model that we state, without proof, that these vectors can be eliminated from (50), yielding the following loss function

$$f(\mathbf{Q},\mathbf{S},\mathbf{W},\mathbf{Y}) = \sum_{j=1}^{M} \text{tr}\,(\mathbf{X}_j - \mathbf{Y}\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j')'\mathbf{C}_j(\mathbf{X}_j - \mathbf{Y}\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j'). \tag{51}$$

In the next subsection, it is first shown that the minimization of (51) with respect to the orthonormal matrix $\mathbf{Q}_j$, dimension weights $\mathbf{W}_j$, and orthonormal matrix $\mathbf{S}_j$ can be reduced to the much more simple problem of minimizing the loss function with respect to one set of parameters only, from which the optimal $\mathbf{Q}_j$, $\mathbf{W}_j$, and $\mathbf{S}_j$ can be recovered afterwards. Then it is discussed how to obtain updates for the centroid configuration $\mathbf{Y}$. Finally, the non-uniqueness properties of the dimension weighting model with idiosyncratic rotations are discussed, and an analysis of variance is presented.

Reducing the estimation of Q_j, S_j, and W_j to one set of parameters

Letting $\mathbf{B}_j = \mathbf{Q}_j\mathbf{W}_j\mathbf{S}_j'$ of order $m_j{\times}m$, (51) may be written as

$$\mathrm{f}(\mathbf{Y},\mathbf{B}) = \sum_{j=1}^{M} \mathrm{tr}\ (\mathbf{X}_j - \mathbf{Y}\mathbf{B}_j')'\mathbf{C}_j(\mathbf{X}_j - \mathbf{Y}\mathbf{B}_j'). \qquad (52)$$

Matrix $\mathbf{B}_j$ stands for the matrix product of a columnwise orthonormal matrix $\mathbf{Q}_j$, a diagonal matrix $\mathbf{W}_j$, and a square orthonormal matrix $\mathbf{S}_j'$. Since a singular value decomposition can be applied to any matrix, no restrictions have to be imposed on matrix $\mathbf{B}_j$. If, therefore, (unrestricted) matrices $\mathbf{B}_j$ can be determined yielding the global minimum of (52) for fixed $\mathbf{Y}$, and given the singular value decompositions

$$\mathbf{B}_j = \mathbf{K}_j\mathbf{\Lambda}_j\mathbf{L}_j' \quad \text{for } j = 1, \ldots, M, \qquad (53)$$

then the global minimum of (51) with respect to $\mathbf{Q}_j$, $\mathbf{W}_j$, and $\mathbf{S}_j$ is obtained by setting $\mathbf{Q}_j = \mathbf{K}_j$, $\mathbf{W}_j = \mathbf{\Lambda}_j$, and $\mathbf{S}_j = \mathbf{L}_j$ for $j = 1, \ldots, M$.

The reformulation of (51) in (52) considerably simplifies the minimization task: instead of having to determine three sets of parameters in (51) only one set of parameters needs to be determined in (52). Moreover, the minimization of (52) over unrestricted matrix $\mathbf{B}_j$ is a simple multivariate multiple regression problem with the well-known solution

$$\mathbf{B}_j = (\mathbf{X}_j'\mathbf{C}_j\mathbf{Y})(\mathbf{Y}'\mathbf{C}_j\mathbf{Y})^{-1}. \qquad (54)$$

Once optimal matrix $\mathbf{B}_j$ have been obtained from (54), the optimal matrices $\mathbf{S}_j$, $\mathbf{W}_j$, and $\mathbf{Q}_j$ can be recovered using (53).

The centroid configuration

For fixed matrices $\mathbf{B}_j$, we have to minimize

$$f(\mathbf{Y}) = \sum_{j=1}^{M} \text{tr } (\mathbf{A}_j - \mathbf{F}_j\mathbf{Y}\mathbf{B}_j')'(\mathbf{A}_j - \mathbf{F}_j\mathbf{Y}\mathbf{B}_j'), \tag{55}$$

where $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$, and $\mathbf{A}_j = \mathbf{F}_j\mathbf{X}_j$. Letting $\mathbf{a}$ be the $MNm_j{\times}1$ partitioned vector containing the M vectors (vec $\mathbf{A}_j$) on top of one another, letting $\mathbf{D}$ be the $MNm_j{\times}Nm$ partitioned matrix containing the M matrices $(\mathbf{B}_j \otimes \mathbf{F}_j)$, and defining $\mathbf{y}$ = (vec $\mathbf{Y}$), the problem of minimizing (55) can be expressed as the classical univariate multiple regression problem:

$$f(\mathbf{y}) = (\mathbf{a} - \mathbf{D}\mathbf{y})'(\mathbf{a} - \mathbf{D}\mathbf{y}). \tag{56}$$

Hence, the global minimum of (55) is attained for

$$\text{vec } \mathbf{Y} = (\mathbf{D}'\mathbf{D})^{-}\mathbf{D}'\mathbf{a} = [\sum_{j=1}^{M} (\mathbf{B}_j'\mathbf{B}_j \otimes \mathbf{C}_j)]^{-}[\text{vec } \sum_{j=1}^{M} \mathbf{C}_j\mathbf{X}_j\mathbf{B}_j], \tag{57}$$

where $[\sum_j (\mathbf{B}_j'\mathbf{B}_j \otimes \mathbf{C}_j)]^{-}$ is the Moore-Penrose inverse of $\sum_j (\mathbf{B}_j'\mathbf{B}_j \otimes \mathbf{C}_j)$. This latter matrix of order $Nm{\times}Nm$ is no longer block-diagonal, and the Moore-Penrose inverse in (57) must therefore be determined for the complete matrix.

If necessary, the computation of the latter Moore-Penrose inverse can be avoided by applying the dimensionwise approach to the updating of $\mathbf{Y}$ discussed in Heiser and Stoop (1986) to (51). Specifically, letting $\mathbf{y}_a$ be the a-th column of $\mathbf{Y}$ ($a = 1,\ldots,m$), $\mathbf{e}_a$ be the a-th column of the identity matrix $\mathbf{I}_m$, and $\mathbf{P}_a$ be the $N{\times}m$ matrix equal to $\mathbf{Y}$ but with the a-th column containing zeroes, then it is true that

$$\mathbf{Y} = \mathbf{P}_a + \mathbf{y}_a\mathbf{e}_a'. \tag{58}$$

Substitution of (58) in (51) yields

$$f(\mathbf{y}_a) = \sum_{j=1}^{M} \text{tr } (\mathbf{X}_j - (\mathbf{P}_a + \mathbf{y}_a\mathbf{e}_a')\mathbf{B}_j')'\mathbf{C}_j(\mathbf{X}_j - (\mathbf{P}_a + \mathbf{y}_a\mathbf{e}_a')\mathbf{B}_j')$$

$$= \sum_{j=1}^{M} \text{tr } (\mathbf{A}_j - \mathbf{y}_a\mathbf{e}_a'\mathbf{B}_j')'\mathbf{C}_j(\mathbf{A}_j - \mathbf{y}_a\mathbf{e}_a'\mathbf{B}_j'), \tag{59}$$

with $\mathbf{A}_j = \mathbf{X}_j - \mathbf{P}_a\mathbf{B}_j'$. Therefore,

$$\text{f}(\mathbf{y}_a) = \text{c}_a + \mathbf{y}_a'(\sum_{j=1}^{M} \mathbf{e}_a'\mathbf{B}_j'\mathbf{B}_j\mathbf{e}_a\mathbf{C}_j)\mathbf{y}_a - 2\,\mathbf{y}_a'(\sum_{j=1}^{M} \mathbf{C}_j\mathbf{A}_j\mathbf{B}_j)\mathbf{e}_a, \tag{60}$$

where c_a is a term independent of $\mathbf{y}_a$. The global minimum of (60) is attained where

$$\mathbf{y}_a = (\sum_{j=1}^{M} \mathbf{e}_a'\mathbf{B}_j'\mathbf{B}_j\mathbf{e}_a\mathbf{C}_j)^{-}(\sum_{j=1}^{M} \mathbf{C}_j\mathbf{A}_j\mathbf{B}_j)\mathbf{e}_a, \tag{61}$$

which only requires the computation of the Moore-Penrose inverse of an $N\times N$ matrix.

In the unweighted case, and assuming that the configurations have been centered on the origin, (57) simplifies to

$$\mathbf{Y} = (\sum_{j=1}^{M} \mathbf{X}_j\mathbf{B}_j)(\sum_{j=1}^{M} \mathbf{B}_j'\mathbf{B}_j)^{-1}, \tag{62}$$

and we only need to determine the proper inverse of an $m\times m$ matrix.

In fact, a much stronger result applies in the unweighted case: the problem of fitting the second dimension weighting model of Lingoes and Borg (1978) to M configurations can then be solved analytically, as follows. Since (54) then can be written as

$$\mathbf{B}_j = \mathbf{X}_j'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}, \tag{63}$$

substitution of (63) in (51) yields

$$\text{f}(\mathbf{Y}) = \text{d} - \sum_{j=1}^{M} \text{tr } \mathbf{X}_j'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{X}_j, \tag{64}$$

where d is independent of $\mathbf{Y}$. Thus, the minimization of (64) is equivalent to the maximization of

$$\text{g}(\mathbf{Y}) = \sum_{j=1}^{M} \text{tr } \mathbf{X}_j'\mathbf{Y}(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{X}_j, \tag{65}$$

Moreover, since we are free to require that $\mathbf{Y'Y} = \mathbf{I}$ in the case of equal column orders (see below), for any $\mathbf{Y}$ satisfying the latter constraint (65) may be written as

$$g(\mathbf{Y}) = \text{tr } \mathbf{Y'AY}, \tag{66}$$

with $\mathbf{A} = \sum_j \mathbf{X}_j\mathbf{X}_j'$. The maximization of (66) subject to $\mathbf{Y'Y} = \mathbf{I}$ is a well-known eigenvalue problem which is solved by setting $\mathbf{Y}$ equal to the first m eigenvectors of matrix $\mathbf{K}$ in the eigenvalue decomposition $\mathbf{A} = \mathbf{K\Lambda K'}$. Once an $\mathbf{Y}$ has thus been obtained, the $\mathbf{B}_j$'s can be computed from (63), after which the singular value decompositions $\mathbf{B}_j = \mathbf{Q}_j\mathbf{W}_j\mathbf{S}_j'$ automatically result in the solution of (53), yielding the global minimum of (51). This is consistent with the IDIOSCAL model for the analysis of (dis)similarity matrices, which can also be solved analytically (Carroll, 1998, personal communication).

Because (52) can be written in the unweighted case as

$$f(\mathbf{Y},\mathbf{B}) = \text{tr } (\mathbf{X} - \mathbf{YB})'(\mathbf{X} - \mathbf{YB}). \tag{67}$$

where $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M]$ and $\mathbf{B} = [\mathbf{B}_1', \mathbf{B}_2', \ldots, \mathbf{B}_M']$, the (analytical) solution of the dimension weighting model with idiosyncratic rotations is, in fact, a principal component analysis of the partitioned $N \times \sum_j m_j$ matrix $\mathbf{X}$. This requires the eigenvalue decomposition of $\mathbf{XX'} = \mathbf{K\Lambda K'}$, which can be written as $\mathbf{A} = \sum_j \mathbf{X}_j\mathbf{X}_j' = \mathbf{K\Lambda K'}$ in (66).

Summarizing, in the general situation of a weighted analysis, a convergent algorithm for the estimation of the unknown parameters in (51) is obtained by alternatingly computing a new (possibly dimensionwise) update for the centroid configuration $\mathbf{Y}$ in one step, and new updates for free matrices $\mathbf{B}_j$ in the second step, until convergence. If the analysis is unweighted, then the fitting of the dimension weighting model with idiosyncratic rotations can be solved analytically through a principal components analysis of matrix $\mathbf{X}$ defined in (67), followed by the computation of (63) and (53) for the determination of the remaining parameter sets $\mathbf{S}$, $\mathbf{W}$, and $\mathbf{Q}$.

Non-uniqueness of the solution and analysis of variance

The value of the loss function for dimension weighting with idiosyncratic rotations is unchanged by the following transformation:

$$f(\mathbf{B},\mathbf{Y}) = \sum_{j=1}^{M} \text{tr}\,(\mathbf{X}_j - \bar{\mathbf{Y}}\bar{\mathbf{B}}_j')'\mathbf{C}_j(\mathbf{X}_j - \bar{\mathbf{Y}}\bar{\mathbf{B}}_j'), \tag{68}$$

where $\bar{\mathbf{Y}} = \mathbf{Y}\mathbf{T}^{-1}$, $\bar{\mathbf{B}}_j = \mathbf{B}_j\mathbf{T}'$, and $\mathbf{T}$ is an arbitrary non-singular matrix of order $m \times m$. An attractive solution is to set $\mathbf{T}$ in (68) equal to

$$\mathbf{T} = \mathbf{G}\mathbf{H}', \tag{69}$$

where $\mathbf{G}$ contains the singular values, and $\mathbf{H}$ the right singular vectors of the singular value decomposition

$$\mathbf{Y} = \mathbf{F}\mathbf{G}\mathbf{H}'. \tag{70}$$

This yields a solution where

$$\bar{\mathbf{Y}} = \mathbf{Y}\mathbf{T}^{-1} = \mathbf{F}\mathbf{G}\mathbf{H}'(\mathbf{G}\mathbf{H}')^{-1} = \mathbf{F}, \tag{71}$$

and

$$\bar{\mathbf{B}}_j = \mathbf{B}_j\mathbf{T}' = \mathbf{B}_j\mathbf{H}\mathbf{G} \tag{72}$$

for $j = 1, \ldots, M$. It follows from (71) that the transformed centroid configuration $\bar{\mathbf{Y}}$ is columnwise orthonormal. Evidently, having defined (72) the singular value decompositions

$$\bar{\mathbf{B}}_j = \mathbf{Q}_j\mathbf{W}_j\mathbf{S}_j' \qquad \text{for } j = 1, \ldots, M \tag{73}$$

are used instead of (53) to obtain the optimal $\mathbf{S}_j$, $\mathbf{W}_j$, and $\mathbf{Q}_j$ in loss function (51).

It follows from (71), (72), and (73) that matrix product $\mathbf{Y}\mathbf{B}_j'$ can be written as

$$\mathbf{Y}\mathbf{B}_j' = \mathbf{V}_j\mathbf{W}_j\mathbf{Q}_j', \qquad \text{for } j = 1, \ldots, M. \tag{74}$$

with $\mathbf{V}_j = \bar{\mathbf{Y}}\mathbf{S}_j = \mathbf{F}\mathbf{S}_j$. Since $\mathbf{V}_j$ is an orthonormal matrix, being the product of orthonormal matrices $\mathbf{F}$ and $\mathbf{S}_j$, (74) is a singular value decomposition of $\mathbf{Y}\mathbf{B}_j'$. Therefore, the matrices $\mathbf{W}_j$

and $\mathbf{Q}_j$ defined by (73) have the nice property that they are unique up to reflections. The same holds for the matrix products $\bar{\mathbf{Y}}\mathbf{S}_j = \mathbf{F}\mathbf{S}_j$.

Only one indeterminacy remains: the matrices $\bar{\mathbf{Y}} = \mathbf{F}$ and $\mathbf{S}_j$ themselves are only determined up to a rotation, because we are still free to write $\mathbf{Y}\mathbf{B}_j'$ as

$$\mathbf{Y}\mathbf{B}_j' = \bar{\mathbf{Y}}\mathbf{P}'\mathbf{P}\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j', \qquad \text{for } j = 1, \ldots, M, \tag{75}$$

where $\mathbf{P}$ is an arbitrary orthonormal matrix of order $m \times m$. In other words, the matrix product $\bar{\mathbf{Y}}\mathbf{P}'$ just defines another orthonormal basis of the raw configuration $\mathbf{Y}$. The interpretation of the present dimension weighting model does not require a unique orientation of the centroid configuration, however, because it is the dimensions of the $\bar{\mathbf{Y}}\mathbf{S}_j$ that are differentially being weighted, and the above procedure guarantees the uniqueness of the latter matrix products.

Another advantage of the choice for (69) is that, if all configurations are complete, just as in the previous dimension weighting model the term $(\mathbf{1}'\mathbf{W}_j^2\mathbf{1})/(\text{tr } \mathbf{X}_j'\mathbf{X}_j)$ is equal to the squared correlation between the elements of $\mathbf{X}_j$ and the elements of $\bar{\mathbf{Y}}\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j'$ (see Commandeur, 1991). Furthermore, whether the objects are weighted or not, and letting $\mathbf{F}_j = \mathbf{N}_j^{1/2}\mathbf{V}_j$, it is always true that

$$r^2(\mathbf{F}_j\mathbf{X}_j, \mathbf{F}_j\bar{\mathbf{Y}}\mathbf{S}_j\mathbf{W}_j\mathbf{Q}_j') = \frac{\text{tr } \mathbf{W}_j\bar{\mathbf{Y}}'\mathbf{C}_j\bar{\mathbf{Y}}\mathbf{W}_j}{\text{tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j}, \tag{76}$$

allowing for an assesment of the fit of each configuration j in the solution.

The analysis of variance of the residuals for the dimension weighting model with idiosyncratic rotations is simply obtained by replacing $\mathbf{Y}$ with $\mathbf{Y}\mathbf{S}_j$ in (49), and everywhere else in the discussion of the analysis of variance for the previous dimension weighting model (see section 4).

6. Illustration: The milk data

Judgments were collected from 10 subjects on 22 types of milk using free choice profiling (Williams and Langron, 1984; Arnold and Williams, 1985). Using rating scales ranging from 0 to 100, the subjects were asked to judge the stimuli on a number of attributes that they had generated themselves. Thus, the 22 types of milk were not only judged on different properties by each subject, but also on different numbers of properties. For the ten subjects these numbers were 31, 29, 31, 29, 31, 24, 33, 30, 38, and 33, respectively. Seven rows of each of the thus obtained data matrices $\mathbf{X}_j$ were randomly selected to be treated as missing in the analysis, and the corresponding weights were set to zero. Throughout the example, and without loss of generality, the normalization $\sum_j \text{tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j = M$ has been applied to the data, and in all algorithms a convergence criterion of 1E-7 has been used.

First, the downward rotation approach of loss function (2) was used to perform a weighted generalized Procrustes analysis of the milk data in two dimensions with the additional estimation of optimal isotropic scaling factors. This yields a solution with the following decomposition of the total variance in the data (see also formula (1)):

$$
\begin{array}{ll}
(1/M) \text{ tr } \mathbf{Z}'\mathbf{C}\mathbf{Z} & = 0.0187 \\
(1/M) \sum_j \text{tr } (s_j\mathbf{X}_j\mathbf{R}_j - \mathbf{Z})'\mathbf{C}_j(s_j\mathbf{X}_j\mathbf{R}_j - \mathbf{Z}) & = 0.0000 \\
(1/M) \sum_j (\text{tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j - s_j^2 \text{ tr } \mathbf{R}_j'\mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j\mathbf{R}_j) & = 0.9813 \\
\hline
(1/M) \sum_j \text{tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j & = 1.0000
\end{array} +
$$

This solution may seem perfect, but is actually useless because almost all of the variance (98 percent) has been rotated out of m-dimensional subspace into the non-fitted dimensions.

INSERT FIGURE 1 ABOUT HERE.

On the other hand, a weighted generalized Procrustes analysis based on the upward rotation method of least-squares criterion (9) yields a solution with the following decomposition (see formula (29)):

$$
\begin{array}{ll}
(1/M) \text{ tr } \mathbf{Z}'\mathbf{C}\mathbf{Z} & = 0.4370 \\
(1/M) \sum_j \text{tr } (s_j\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j')'\mathbf{C}_j(s_j\mathbf{X}_j - \mathbf{Z}\mathbf{R}_j') & = 0.5630 \\
\hline
(1/M) \sum_j \text{tr } \mathbf{X}_j'\mathbf{C}_j\mathbf{X}_j & = 1.0000
\end{array} +
$$

showing that 44 percent of the variance is explained by the two-dimensional solution. Figure 1 contains plots of $\bar{\mathbf{Z}} = \mathbf{ZK}$ (that is, $\mathbf{Z}$ rotated to principal components) obtained with the latter analysis, labeled by two objective characteristics of the 22 types of milk under investigation: method of conservation, and amount of fat. As can be seen in Figure 1, the pasteurized and the sterilized milks are perfectly separated on the line with an angle of about 45 degrees with the horizontal axis. On the other hand, although this distinction is not as clear cut as the one concerning method of conservation, the line with an angle of about minus 45 degrees with the horizontal axis distinguishes between the milks in terms of the amount of fat they contain.

INSERT FIGURE 2 ABOUT HERE.

INSERT FIGURE 3 ABOUT HERE.

In Figures 2 and 3 the idiosyncratic attributes of judges 1, 4, 6, and 7 have been plotted as vectors in the space of the 22 types of milk. To this end the values in the matrices $\bar{\mathbf{R}}_j = \mathbf{R}_j\mathbf{K}$ (see section 3) are used as coordinates for the corresponding attribute vectors. To keep the plots readable only the names of the attributes corresponding to longer vectors are shown. The reason that some attributes are represented more than once in some plots is that the attributes were generated from the subjects using a number of general categories (smell, taste, sight, etc.). In this way, it is possible to generate identical attributes of the stimuli for two or more of these general categories from a subject. Superimposing the plots in Figure 1 of the 22 types of milk on top of the plots in Figures 2 and 3, the thus obtained biplots show how attributes like "cooked taste", "bitter", and "burnt" point towards sterilized milks and away from pasteurized milks, while attributes like "(translucent) watery" and "white blue/green" point towards skim milks, and "fatty", "creamy", and "sticky" in the direction of whole milks.

INSERT FIGURE 4 ABOUT HERE.

The milk data were also subjected to the dimension weighting analysis (without idiosyncratic rotations) described in section 4. This yields a solution which explains half of the total variance in the data, six percent more than the GPA solution. The centroid configuration for this analysis is shown in Figures 4, again labeled with method of conservation and amount of fat. The interpretation of this configuration is identical to the one obtained in GPA; however, the interpreted axes are now the horizontal and the vertical axis, respectively, showing at least for this example the "unique axes" property of INDSCAL type of models.

INSERT FIGURE 5 ABOUT HERE.

Figure 5 contains a plot of the dimension weights of the ten subjects. The plot shows that the second subject puts much larger emphasis on the method of conservation distinction than on the amount of fat differentiation, while this situation is reversed for the sixth subject. The remaining subjects are located somewhere in between these two extremes. Letting $\mathbf{A}_j = \mathbf{YW}_j$ denote the individual space of a particular subject, it may be noted that the matrix products $\mathbf{A}_j\mathbf{Q}_j'$ again define m-dimensional linear biplots in which the original attributes $\mathbf{Q}_j$ of subject j can be displayed as vectors, together with the stimulus points in his or her individual space $\mathbf{A}_j$.

A dimension weighting analysis of the milk data including idiosyncratic rotations resulted in a solution which explains 51 percent of the total variance of the data. Since the improvement over the previous analysis is only one percent, we do not further discuss the results of this analysis.

7. Conclusions

A new approach to the matching of configurations in subspaces of lower dimensionality has been presented. The method is least-squares, and replaces downward rotations of the high-dimensional data matrices with upward rotations of the low-dimensional centroid configuration. Even though being least-squares, the latter method optimizes the "consensus" criterion proposed by Peay (1988). In contrast with downward rotations, the upward rotation method preserves relative distances in generalized Procrustes analysis, is not sensitive to rotating out variance, and is computationally efficient. When analyses are performed in the full space of configurations containing an equal number of columns, applying orthonormal transformations to the centroid configuration comes down to the same thing as orthonormally transforming the given configurations, and the solutions are identical.

The method of upward rotations has been applied to three different models for the matching of configurations allowing for increasingly more complex transformations: the GPA model, and two dimension weighting models that are related to INDSCAL and IDIOSCAL in terms of interpretation. The possibility of performing weighted analyses is included in all three models, of which the matching of matrices with missing rows is a special case.

We end up by noting that the method proposed in the present paper also compares favorably with the method of "padding out with zeroes" proposed by Dijksterhuis and Gower (1992) and Gower (1995). They advised to append zero columns to the configurations such that they all end up having $p = \max\{m_j\}$ columns, and then performing the analysis in the full p-dimensional space. Although the principal m dimensions of the p-dimensional centroid configuration obtained with this method seem to be quite similar to the m-dimensional solution obtained with the upward rotation method proposed in the present paper, the latter solution explains more variance in the data. For example, a 38-dimensional generalized Procrustes analysis of the milk data discussed in the previous section using the method of "padding out with zeroes" results in a solution whose two principal dimensions explain 39.6 percent of the total variance, while the two-dimensional subspace solution explains 43.7 percent of the variance. Finally, "padding out with zeroes" requires the computation of p-dimensional rotation matrices during iterations, while upward rotations are only of size m.

References

Arnold, G.M., & Williams, A.A. (1985). The use of generalised Procrustes analysis in sensory analysis. In J.R. Piggott (Ed.), *Statistical procedures in food research*. Amsterdam: North Holland.

Borg, I., & Groenen, J. (1997). *Modern multidimensional scaling: theory and applications.* New York: Springer-Verlag.

Carroll, J.D., & 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., & Chang, J.J. (1972). *IDIOSCAL (Individual Differences In Orientation SCALing): A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL.* Paper presented at the Psychometric Society Meeting, Princeton, N.J.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, *31*, 33 - 42.

Commandeur, J.J.F. (1991). *Matching configurations*. Leiden: DSWO Press.

De Leeuw, J. & Meulman, J. (1986). A special Jackknife for Multidimensional Scaling. *Journal of Classification*, *3*, 97-112.

Dijksterhuis, G.B., & Gower, J.C. (1992). The interpretation of generalised Procrustes analysis and allied methods. *Food Quality and Preference*, *3*, 67-87.

Everitt, B.S., & Gower, J.C. (1981). Plotting the optimum positions of an array of cortical electrical phosphenes. In V.D. Barnett (Ed.), *Interpreting multivariate data* (pp.279-287). Chichester: Wiley.

Gower, J.C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*, 33 -51.

Gower, J.C. (1995). Orthogonal and projection Procrustes analysis. In W.J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp.113-134). Oxford: Clarendon Press.

Green, B.F., & Gower, J.C. (1979). *A problem with congruence*. Paper presented at the Annual Meeting of the Psychometric Society, Monterey, California.

Groenen, P.J.F., Commandeur, J.J.F. & Meulman, J.J. (1996). *Distance analysis of large data sets of categorical variables using object weights*. Research Report RR-96-01. Leiden: Department of Data Theory.

Heiser, W.J., & Stoop, I. (1986). *Explicit SMACOF algorithms for individual differences scaling*. Research report RR-86-04. Leiden: Department of Data Theory.

Koschat, M.A., & Swayne, D.F. (1991). A weighted Procrustes criterion. *Psychometrika*, *56*, 229-239.

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

Meulman, J.J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika*, *57*, 539-565.

Meulman, J.J. (1998). A distance-based biplot for multidimensional scaling of multivariate data. In C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H.-H. Boak, & Y. Baba (Eds.), *Data science, classification, and related methods* (pp. 506-517). Springer-Verlag: Tokyo.

Mooijaart, A., & Commandeur, J.J.F. (1990). A general solution of the weighted orthonormal Procrustes problem. *Psychometrika*, *55*, 657-663.

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

Schönemann, P.H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*, 1 - 10.

ten Berge, J.M.F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, *42*, 267 - 276.

ten Berge, J.M.F., & Bekker, P.A. (1993). The isotropic scaling problem in Generalized Procrustes Analysis. *Computational Statistics & Data Analysis*, *16*, 201-204.

ten Berge, J.M.F., & Knol, D.L. (1984). Orthogonal rotations to maximal agreement for two or more matrices of different column orders. *Psychometrika*, *49*, 49 - 55.

ten Berge, J.M.F., Kiers, H.A.L., & Commandeur, J.J.F. (1993). Orthogonal Procrustes rotation for matrices with missing values. *British Journal of Mathematical and Statistical Psychology*, *46*, 119-134.

Williams, A.A., & Langron, S.P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the science of food and agriculture*, *39*, 558-568.

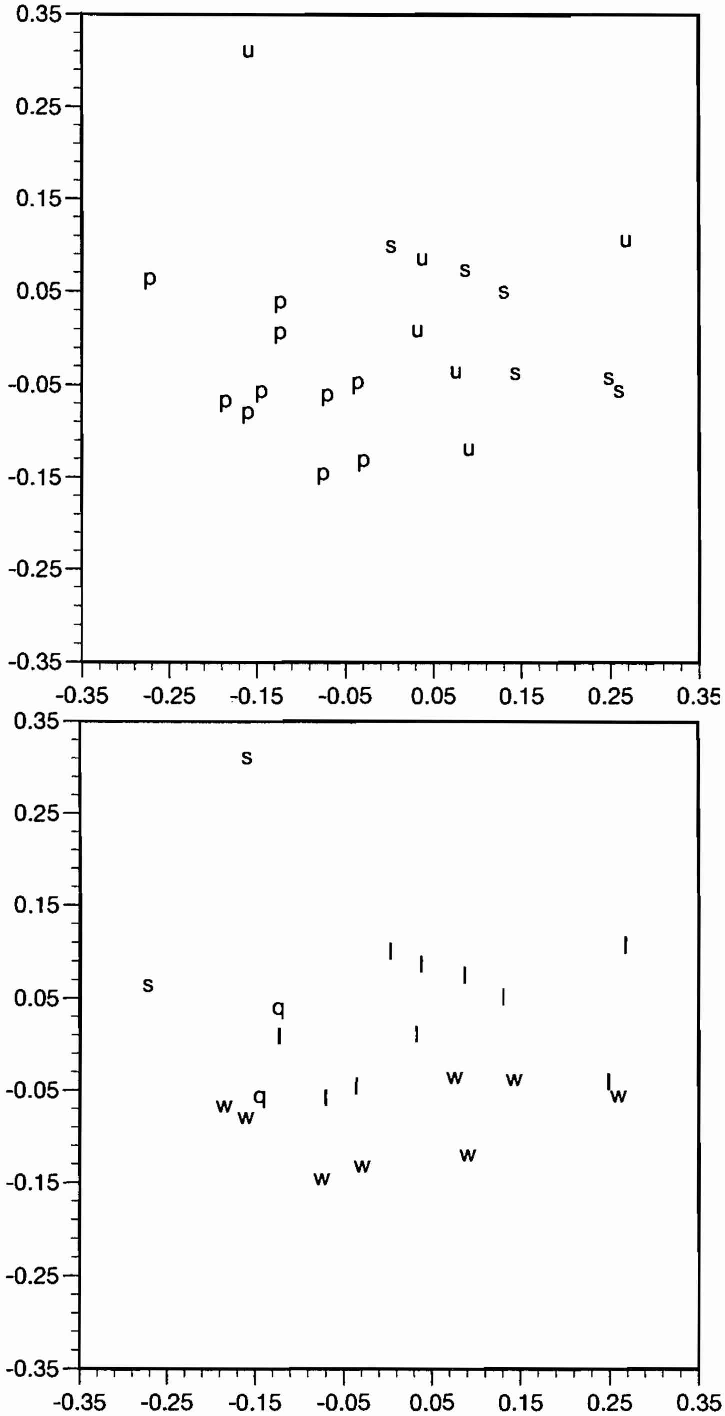


Figure 1. Two-dimensional centroid configuration obtained with a weighted GPA of 22 types of milk using upward rotations, labeled by conservation method (top: p = pasteurization, s = sterilization, u = ultrahigh temperature sterilization), and amount of fat (bottom: s = skim milk, q = quarter fat, l = low fat, w = whole milk), fit = 0.44.

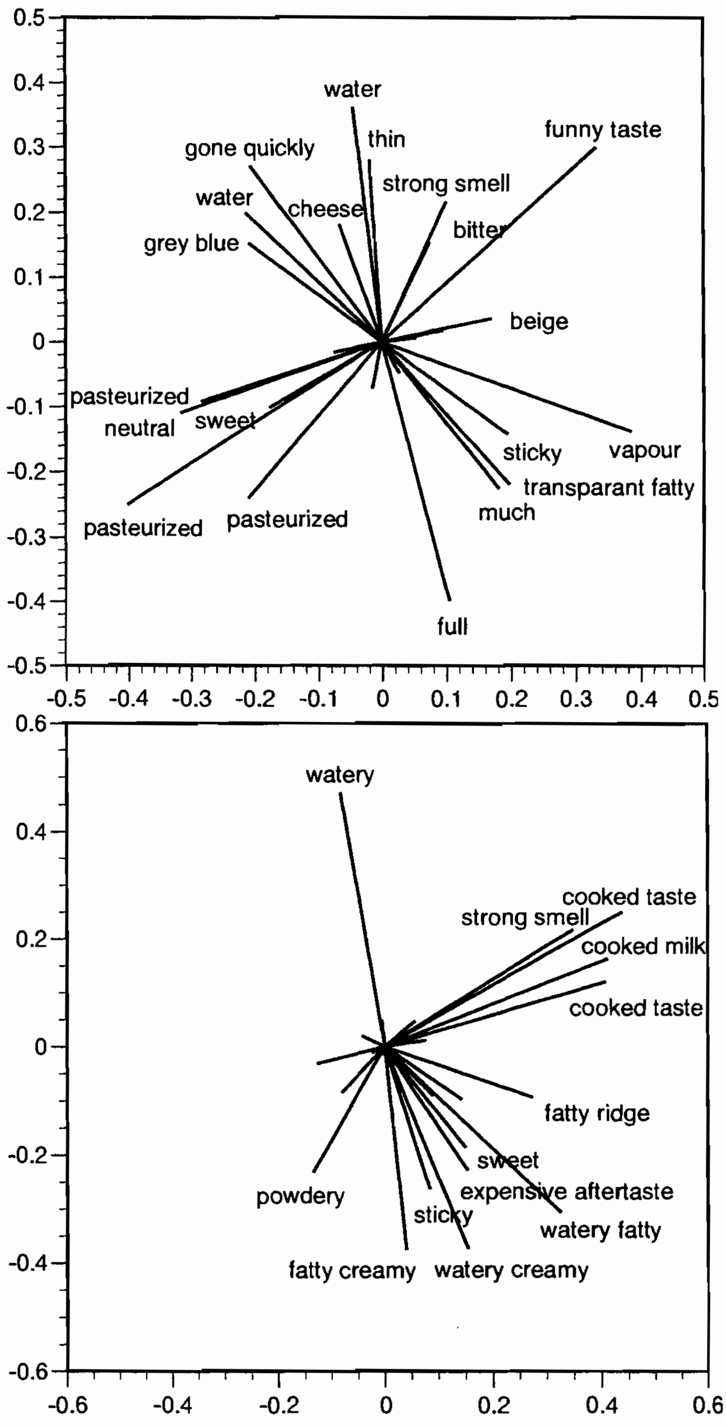


Figure 2. GPA plot of 31 attributes of judge 1 (top), and 29 attributes of judge 4 (bottom), combining with the centroid configuration in Figure 1 to a linear biplot of the attributes and stimuli under investigation.

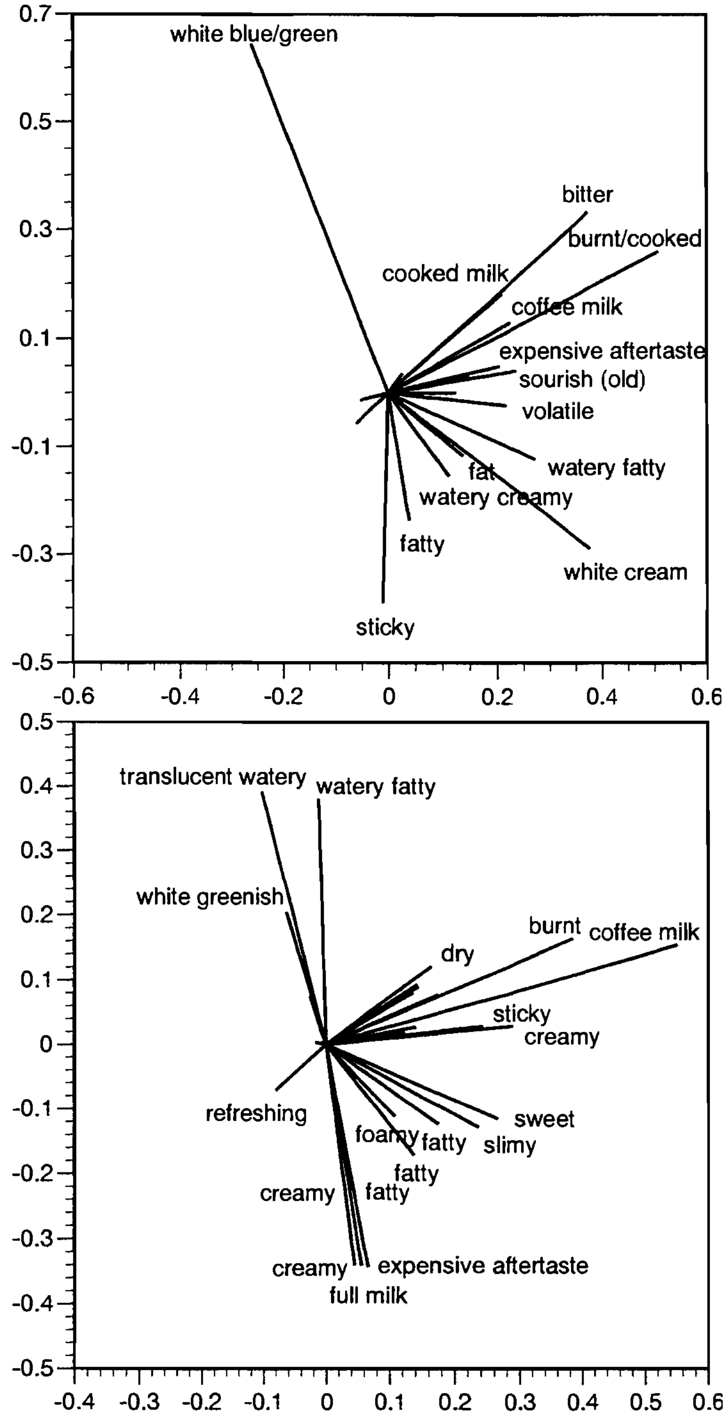
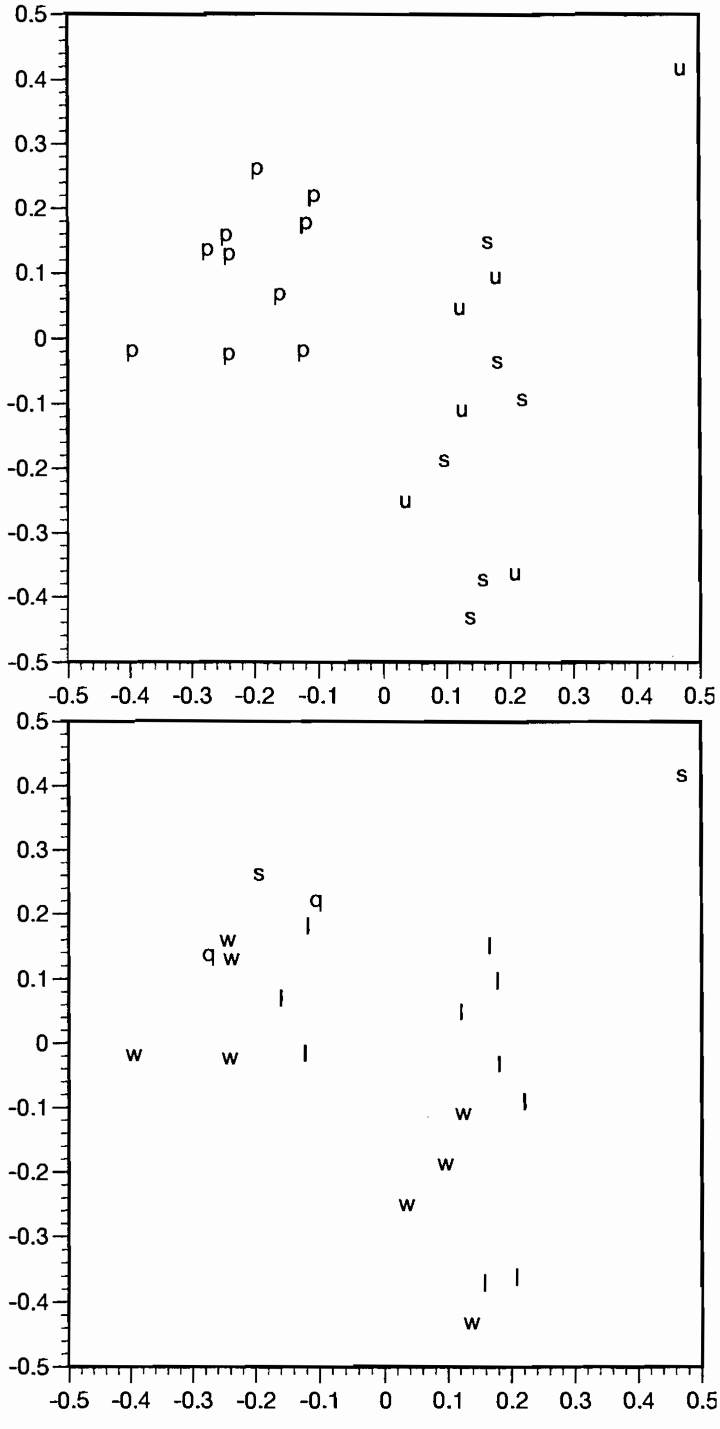


Figure 3. GPA plot of 24 attributes of judge 6 (top), and 33 attributes of judge 7 (bottom), combining with the centroid configuration in Figure 1 to a linear biplot of the attributes and stimuli under investigation.



<u>Figure 4.</u> Two-dimensional centroid configuration of a weighted dimension weighting analysis of 22 types of milk using upward rotations, labeled by conservation method (top: p = pasteurization, s = sterilization, u = ultrahigh temperature sterilization), and amount of fat (bottom: s = skim milk, q = quarter fat, l = low fat, w = whole milk), fit = 0.50.

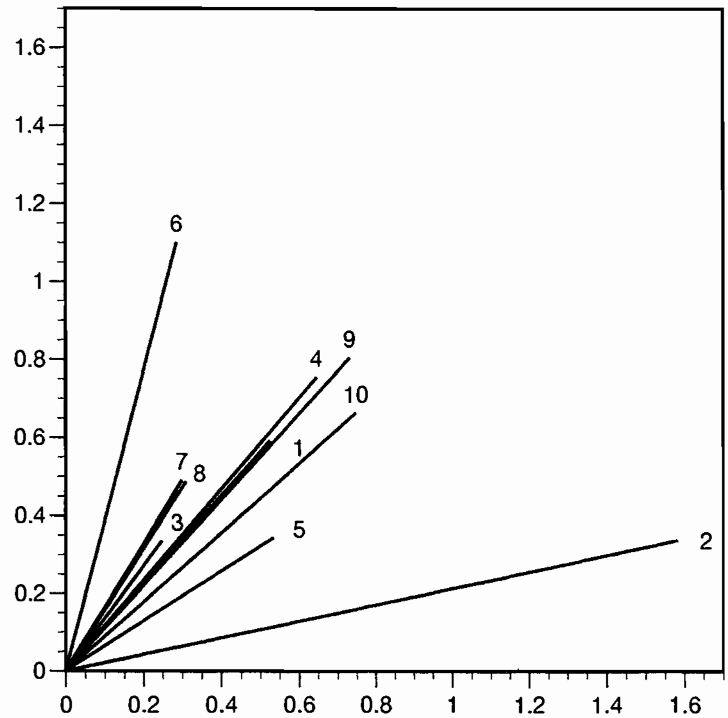


Figure 5. Two-dimensional subject space of the weighted dimension weighting analysis of 22 types of milk using upward rotations, labeled by subject number.