RR 88-01

# REDUCED SPACE ANALYSIS

JACQUELINE MEULMAN

DEPARTMENT OF
DATA THEORY

# REDUCED SPACE ANALYSIS

A particular form of generalized canonical analysis compares $p$-dimensional linear combinations of $M$ sets of variables with a $p$-dimensional unknown matrix $\mathbf{X}$. This technique can be shown to have certain distance properties in terms of the objects in the analysis. From this starting point various modifications of the technique can be introduced that optimize these distance properties. In this paper a quite general approach is presented that has a previous generalization (Meulman, 1986) as a special case.

## 1. INTRODUCTION

Generalized canonical analysis is the extension of canonical correlation analysis to the analysis of more than two sets of variables. The history of the technique shows that there are different ways to do this, based on a different criterion in terms of the correlation matrix between the canonical variables. A canonical variable is a linear combination of the original variables in one of the sets. For each of the $M$ sets we look for $p$ canonical variables. We need some constraints in order to secure canonical variables from the same set to be different. Two types of constraints can be distinguished, called the strong and the weak orthogonality constraints by Dauxois and Pousse (1976). Unifying discussions on the various approaches to generalized canonical analysis can be found a.o. in Kettenring (1971), Gifi (1981), and Van de Geer (1984, 1986). In this paper a particular generalization will be taken as starting point. It has been discussed by Carroll (1968), Saporta (1975), Gifi (1981), De Leeuw (1984), Van der Burg and De Leeuw (1987) and Van der Burg, De Leeuw and Verdegaal (1984, 1986).

In this approach each $p$-dimensional set of canonical variables is compared with an unknown configuration $\mathbf{X}$. This comparison matrix gives the coordinates of the objects in a $p$-dimensional space. The sets of original variables are denoted by $\mathbf{H}_J$ ($J$=1,...,$M$), defined on the same $n$ objects. In the set $\mathbf{H}_J$ we have $m_J$ variables. For each set we have to find an $m_J$x$p$ weight matrix $\mathbf{A}_J$. Now the loss function for this problem can be written as

$$\sigma(\mathbf{A},\mathbf{X}) = M^{-1} \sum_{J=1}^{M} \operatorname{tr} (\mathbf{H}_J\mathbf{A}_J - \mathbf{X})'(\mathbf{H}_J\mathbf{A}_J - \mathbf{X}), \tag{1}$$

which is to be minimized over the comparison space $\mathbf{X}$ and over the weights $\mathbf{A}_J$ for each of the $M$ sets. One way to minimize this loss function is by alternating least squares: in the first step we compute the weights for fixed $\mathbf{X}$ and in the second step we solve for $\mathbf{X}$ for given $\mathbf{A}_J$. Then we improve the weights again, and continue until convergence is attained. In one of the two steps we have to impose the constraints mentioned above. A very convenient constraint is to require that $\mathbf{X}'\mathbf{X} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. With respect to the $\mathbf{H}_J\mathbf{A}_J$ it implies that this is equivalent to the weak orthogonality constraints, written as $M^{-1}\Sigma_J\mathbf{A}_J'\mathbf{H}_J'\mathbf{H}_J\mathbf{A}_J = \mathbf{I}$. The strong orthogonality constraints would require that each $\mathbf{A}_J'\mathbf{H}_J'\mathbf{H}_J\mathbf{A}_J = \mathbf{I}$. It can be shown that the minimum loss of (1) is a function of the eigenvalues of a particular scalar product matrix, written as $M^{-1}\Sigma_J\mathbf{H}_J(\mathbf{H}_J'\mathbf{H}_J)^{-1}\mathbf{H}_J'$. The configuration $\mathbf{X}$ turns out to be equal to the first $p$ eigenvectors of this matrix.

Loss function (1) has been modified to incorporate optimal nonlinear transformations $\mathbf{q}_j = \phi_j(\mathbf{h}_j)$, where $\phi_j$ ($j$=1,...,$m_J$) denotes a nonlinear function for variable $j$ in set $J$ (cf. Gifi, 1981; Van der Burg, De Leeuw and Verdegaal, 1986; see Van der Burg and Dijksterhuis, 1988, for an application in the multiway data analysis context). The class of nonlinear transformations includes nominal and ordinal transformations, and linear transformations are a special case.

## 2. DISTANCE PROPERTIES OF CANONICAL ANALYSIS

In Meulman (1986) it is shown in line with Gower (1966) that the form of generalized canonical analysis described in section 1 has certain distance properties. The optimal solution for the coordinates of the objects in $\mathbf{X}$ solves a particular *multidimensional scaling* problem: the Euclidean distances between the object points approximate the Mahalanobis distances in the high-dimensional spaces spanned by the variables in the $M$ different sets. The squared Euclidean distance matrix between the objects in $\mathbf{X}$ is given by

$$\mathrm{D}^2(\mathbf{X}) = \mathbf{au}' + \mathbf{ua}' - 2\mathbf{XX}' \tag{2}$$

where $\mathbf{a}$ is the $n$-vector containing the diagonal elements of $\mathbf{XX}'$ and $\mathbf{u}$ an $n$-vector with elements equal to 1. The squared Mahalanobis distance matrix between the objects in set $J$ is defined by

$$\mathrm{D}^2_{\mathrm{V}}(\mathbf{H}_J) = \mathbf{b}_J\mathbf{u}' + \mathbf{ub}_J' - 2\mathbf{H}_J\mathbf{V}_{\mathrm{H}}^{-1}\ \mathbf{H}_J' = \mathrm{D}^2(\mathbf{H}_J\mathbf{V}_{\mathrm{H}}^{-1/2}) \tag{3}$$

where $\mathbf{b}_J$ is the $n$-vector containing the diagonal elements of $\mathbf{H}_J\mathbf{V}_{\mathrm{H}}^{-1}\mathbf{H}_J'$, and $\mathbf{V}_{\mathrm{H}}^{-1}$ is the inverse of the covariance matrix $\mathbf{H}_J'\mathbf{H}_J$ (we assume the columns of $\mathbf{H}_J$ in deviation from their means). The Mahalanobis distances $\mathrm{D}_{\mathrm{V}}(\mathbf{H}_J)$ incorporate the covariances between the variables in the same set, and can be interpreted as Euclidean distances $\mathrm{D}(\mathbf{H}_J\mathbf{V}_{\mathrm{H}}^{-1/2})$ in an *orthonormal* space. The coordinates in $\mathbf{X}$ that are obtained when (1) is minimized are the same as would have been obtained when the classical, Torgerson (1958) and Gower (1966) MDS approach is applied. This comes to the same (Meulman ,1986) as minimizing the so-called STRAIN loss function, defined by

$$\mathrm{STRAIN}(\mathbf{X}) = M^{-1}\sum_{J=1}^{M} \mathrm{tr}\ \mathbf{J}(\mathrm{D}^2(\mathbf{H}_J\mathbf{V}_{\mathrm{H}}^{-1/2})-\mathrm{D}^2(\mathbf{X}))\mathbf{J}'(\mathrm{D}^2(\mathbf{H}_J\mathbf{V}_{\mathrm{H}}^{-1/2})-\mathrm{D}^2(\mathbf{X}))\mathbf{J}. \tag{4}$$

Here $\mathbf{J}$ is the centering operator $\mathbf{J} = \mathbf{I}\text{-}\mathbf{uu}'/\mathbf{u}'\mathbf{u}$. Taking the squares of the distances and the double centering operation defines the Young-Householder (1938) process that turns a set of squared distances into a set of scalar products. By partitioning the STRAIN function it turns out that the obtained scalar product matrix equals $M^{-1}\Sigma_J \mathbf{H}_J(\mathbf{H}_J'\mathbf{H}_J)^{-1}\mathbf{H}_J'$. So the eigenvalue decomposition of this matrix in order to approximate it by a scalar product matrix of lower rank, must render eigenvectors that are equal to the optimal $\mathbf{X}$ from (1). Only here we normalize $\mathbf{X}$ not to be orthonormal, merely to be diagonal, with the sum of squares of its columns equal to the eigenvalues.

We give an example of the technique to illustrate certain peculiarities in terms of the distance approximation. We re-analyze data given by Cailliez and Pagès (1976). The variables are different measurements on 24 fish that were placed in three aquariums, contaminated with the same amount of radioactive strontium. Fish 1-8 stayed for a short time in aquarium 1, fish 9-17 stayed for a longer time in aquarium 2, and Fish 18-24 stayed for the longest time in aquarium 3. Fish 17 died during the experiment. We analyze the variables in three different sets (cf. Gifi, 1981), where set 1 contains 9 variables with measures of radioactivity of various body parts of the fish, set 2 contains 7 variables with measurements of the fish, and set 3 indicates the aquarium. The variables in the first two sets were treated numerically, while the variable in the third set was treated multiple nominally.

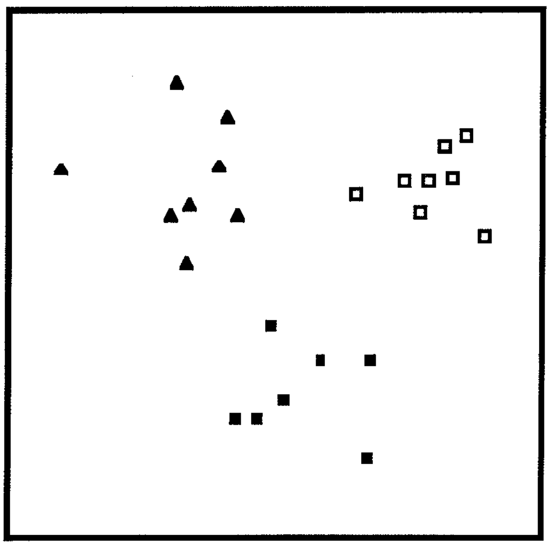


**Figure 1**. Two-dimensional configuration for radioactive fish (◻ shortest exposure, ▲ average exposure, ■ longest exposure)

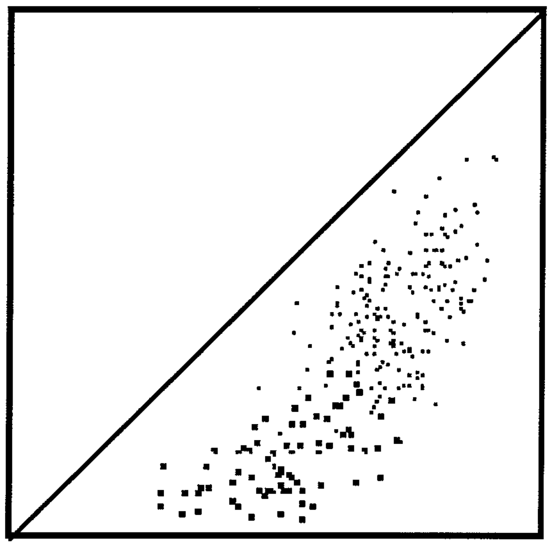


**Figure 2**. Plot of the residuals. Fitted distances (vert. axis) versus average Mahalanobis distances (hor. axis). Approximation from below.

The two-dimensional comparison space **X** is given in Figure 1. The technique recovers the three clusters perfectly: there is no fish that is closer to a fish from another aquarium than it is to a fish from its own aquarium (a □ denotes the fish that were the shortest time exposed, a ▲ the fish that were longer exposed, and a ■ the fish that were thelongest time in the radioactive aquarium). It is remarkable that it is the second dimension (with eigenvalue .760) that discriminates between the fish that were exposed most and the rest. The first dimension (eigenvalue .839) discriminates only between the short and average exposure. On the basis of this configuration the fit of the original variables (in terms of the component loadings as in principal components analysis) was computed, which indicated that the variables of the first set fitted much better on the average (.67) than the variables of the second set (average .38). Especially the measures on radioactivity of eyes, gills, gill covers, fins and scales fit very well; the variable that measures the radioactivity of the kidney has a small fit (.12). Figure 2 displays the residuals in terms of distances. On the vertical axis the fitted distances are plotted versus the average Mahalanobis distances on the horizontal axis. It is clear that all fitted distances are smaller; that is why the approximation is called *from below*. Some scatter points have a bold label ■ ; these are all distances between the fish that have been in the same aquarium (the within-group distances). It appears that not all original within-group distances are smaller than the between-group distances; the larger within-distances are more severely approximated from below.

## 3. OPTIMIZING THE DISTANCE PROPERTIES

Using the distance interpretation of generalized canonical analysis, the Torgerson-Gower approach to MDS can be replaced (Meulman, 1986) by the Kruskal (1964) or Guttman (1968) approach, in its *metric* variant. This amounts to minimizing of the so-called STRESS function

$$\mathrm{STRESS}(\mathbf{X}) = M^{-1} \sum_{J=1}^{M} \mathrm{tr}\,(\mathrm{D}(\mathbf{H}_J \mathbf{V}_{\mathrm{H}}^{-1/2}) - \mathrm{D}(\mathbf{X}))'(\mathrm{D}(\mathbf{H}_J \mathbf{V}_{\mathrm{H}}^{-1/2}) - \mathrm{D}(\mathbf{X})), \tag{5}$$

over de configuration space **X**: each of the $M$ sets of Mahalanobis distances $\mathrm{D}(\mathbf{H}_J \mathbf{V}_{\mathrm{H}}^{-1/2})$ is approximated by the single set of distances D(**X**). The minimization of STRESS gives a solution for the configuration space that improves upon the classical solution in terms of distances, and least squares distance fitting implies that the approximation is from below and from above (cf. section 5).

The minimum of (5) can be found by using an existing multidimensional scaling program, like e.g. KYST (Kruskal, Young and Seery, 1973), or MULTISCALE (Ramsay, 1982), or SMACOF (Heiser and De Leeuw, 1977). These procedures should be applied to the average Mahalanobis distances

$$\underline{D}_V = M^{-1}\sum_{J=1}^{M} D(\mathbf{H}_J\mathbf{V}_H^{-1/2}). \tag{6}$$

This can be seen from the following. STRESS(X) (5) can be partitioned into two additive components, where the first components is written as

$$\text{STRESS}_H(\mathbf{X}) = M^{-1}\sum_{J=1}^{M} \text{tr}\,(D(\mathbf{H}_J\mathbf{V}_H^{-1/2})-\underline{D}_V)'(D(\mathbf{H}_J\mathbf{V}_H^{-1/2})-\underline{D}_V), \tag{7}$$

which measures stress due to heterogeneity and is a constant term, while the second component, the so-called proper stress, defined by

$$\text{STRESS}_P(\mathbf{X}) = \text{tr}\,(\underline{D}_V-D(\mathbf{X}))'(\underline{D}_V-D(\mathbf{X})), \tag{8}$$

is the only component that is dependent on **X**.

As was the case for the ordinary generalized canonical analysis loss function (1), the stress function (5) can also be extended to incorporate optimal transformations of the variables. The purpose of the latter is to improve upon the goodness-of-fit by allowing for nonlinear transformations of the variables that span the high-dimensional space. Because in multidimensional scaling we like the configuration **X** to have a certain shape, we do not normalize **X**. In stead, we have to deal with the strong orthogonality constraints because $\mathbf{V}_Q^{-1/2}\mathbf{Q}_J'\mathbf{Q}_J\mathbf{V}_Q^{-1/2} = \mathbf{I}$ by definition. It might be clear that in this case $\text{STRESS}_H(\mathbf{X})$ is not a constant term anymore, which makes the minimization of STRESS(X) in combination with the constraints much more complicated. How this can be done was shown in Meulman (1986), where also applications are given that compare the optimal solutions of (1) and (5), both with and without optimal transformations of variables.

## 4. DISTANCES IN REDUCED CANONICAL SPACES

Canonical analysis in (5) approximates Mahalanobis distances in $m_J$-dimensional spaces $\mathbf{H}_J$; in (1) it operates on the high-dimensional spaces $\mathbf{H}_J$ by applying the rank reduction $\mathbf{Z}_J = \mathbf{H}_J\mathbf{A}_J$. In this section a technique will be presented that combines features from the two approaches. Incorporating at the same time the optimal transformations $\mathbf{q}_j$ the technique can be formalized by

$$\text{STRESS}(\mathbf{Q},\mathbf{A},\mathbf{X}) = M^{-1}\sum_{J=1}^{M} \text{tr}\,(D(\mathbf{Q}_J\mathbf{A}_J) - D(\mathbf{X}))'(D(\mathbf{Q}_J\mathbf{A}_J) - D(\mathbf{X})). \tag{9}$$

Like in (1) we obtain a reduced space, like in (5) we find a comparison space **X** that is optimal in terms of distances between the objects. A very special feature of (9) is that it is not required that the reduced canonical space $\mathbf{Z}_J = \mathbf{Q}_J\mathbf{A}_J$ has the same dimensionality as **X**, like in (1). In fact, each space $\mathbf{Z}_J$ can have its own number of dimensions, and the latter number, denoted by

$r_J$, can be chosen as $p \le r_J \le m_J$, but also as $r_J \le p \le m_J$. When $r_J$ is chosen as $m_J$, we are back at (5). This gives the new technique a considerable amount of flexibility, which is obtained by defining optimality via the distance function D(.).

It is important to note that we have to put the strong orthogonality constraints on the $\mathbf{Q}_J\mathbf{A}_J$: it must be true that $\mathbf{A}_J'\mathbf{Q}_J'\mathbf{Q}_J\mathbf{A}_J = \mathbf{I}$ in order to deal with Mahalanobis distances. Finding the optimal transformation of the $\mathbf{h}_j$ in combination with this restriction makes the minimization of (9) far from trivial. We will discuss the algorithm in a number of steps. As was shown in section 3 the loss function can be partitioned into additive components. This means that in each step that updates $\mathbf{X}$ we only have to take the average distances between the $\mathbf{Q}_J\mathbf{A}_J$ into account. We define the update step for $\mathbf{X}$ as

$$\hat{\mathbf{X}} = \mathrm{U}(\mathbf{X}) = n^{-1}\mathrm{B}(\mathbf{X})\mathbf{X}, \tag{10}$$

where the B(X) matrix is defined by

$$\begin{array}{lll} \mathrm{B}(\mathbf{X}) & = \mathrm{B}^*(\mathbf{X}) - \mathrm{B}^{o}(\mathbf{X}), & \text{with:} \\ \mathrm{b}_{ik}^{o}(\mathbf{X}) & = M^{-1}\, \Sigma_J\, \mathrm{d}_{ik}(\mathbf{Q}_J\mathbf{A}_J) \,/\, \mathrm{d}_{ik}(\mathbf{X}) & \text{if } i \neq k; \\ \mathrm{b}_{ik}^{o}(\mathbf{X}) & = 0 & \text{if } \mathrm{d}_{ik}(\mathbf{X}) = 0; \\ \mathrm{B}^*(\mathbf{X}) & = \text{diagonal}, & \text{with } \mathrm{b}_{ii}^{*} = \mathbf{u}'\mathrm{B}^{o}(\mathbf{X})\mathbf{e}_i. \end{array} \tag{11}$$

with $\mathbf{e}_i$ the $i$th column of $\mathbf{I}$.

In the next $M$ steps we have to find the updates for the $\mathbf{Q}_J$ and the $\mathbf{A}_J$ The first preliminary step is quite easy: we simply compute a coordinate system $\mathbf{Y}_J$ that improves the fit for given $\mathbf{X}$. This can be done completely analogous to the updating step for $\mathbf{X}$, only now the role of $\mathbf{X}$ and $\mathbf{Q}_J\mathbf{A}_J$ is reversed. So here we compute the update

$$\hat{\mathbf{Y}}_J = \mathrm{U}(\mathbf{Q}_J\mathbf{A}_J) = \mathrm{n}^{-1}\, \mathbb{B}(\mathbf{Q}_J\mathbf{A}_J)\mathbf{Q}_J\mathbf{A}_J \tag{12}$$

with the reversed $\mathbb{B}(\mathbf{Q}_J\mathbf{A}_J)$ matrix defined by

$$\begin{array}{lll} \mathbb{B}(\mathbf{Q}_J\mathbf{A}_J) & = \mathbb{B}^*(\mathbf{X}) - \mathbb{B}^{o}(\mathbf{X}), & \text{with:} \\ \mathbb{b}_{ik}^{o}(\mathbf{Q}_J\mathbf{A}_J) & = \mathrm{d}_{ik}(\mathbf{X}) \,/\mathrm{d}_{ik}(\mathbf{Q}_J\mathbf{A}_J) & \text{if } i \neq k; \\ \mathbb{b}_{ik}^{o}(\mathbf{Q}_J\mathbf{A}_J) & = 0 & \text{if } \mathrm{d}_{ik}(\mathbf{Q}_J\mathbf{A}_J) = 0; \\ \mathbb{B}^*(\mathbf{Q}_J\mathbf{A}_J) & = \text{diagonal}, & \text{with } \mathbb{b}_{ii}^{*} = \mathbf{u}'\mathbb{B}^{o}(\mathbf{Q}_J\mathbf{A}_J)\mathbf{e}_i. \end{array} \tag{13}$$

It will be clear that in general this $\mathbf{Y}_J$ will not meet the analysis requirements: i.e. it should consist of the linear combination $\mathbf{Q}_J\mathbf{A}_J$, while $\mathbf{A}_J'\mathbf{Q}_J'\mathbf{Q}_J\mathbf{A}_J = \mathbf{I}$ and $\mathbf{q}_j \in \mathrm{C}_j$. The cone $\mathrm{C}_j$ is defined by the set of all admissible transformations of variable $\mathbf{h}_j$. Since, however, $\hat{\mathbf{Y}}_J$ is a good candidate in terms of distances, we wish to stay as close to $\hat{\mathbf{Y}}_J$ as possible. So we have to solve the metric projection problem

$$P_C(\mathbf{Y}_J) = \min_{\mathbf{Q}_J,\mathbf{A}} \ \mathrm{tr}\,(\hat{\mathbf{Y}}_J - \mathbf{Q}_J\mathbf{A}_J)'(\hat{\mathbf{Y}}_J - \mathbf{Q}_J\mathbf{A}_J), \quad \text{statisfying } \mathbf{A}_J'\mathbf{Q}_J'\mathbf{Q}_J\mathbf{A}_J = \mathbf{I}, \ \text{and } \mathbf{q}_{j\in J} \in C_j. \tag{14}$$

Suppose we wish to find monotone transformations of the $\mathbf{H}_J$. Then $\mathbf{Q}_J$ should satisfy the ordinality constraints. This is not a simple problem, which is also discussed in Van der Burg and De Leeuw (1983) in relation with nonlinear canonical correlation analysis with strong orthogonality constraints. In their case $\mathbf{Y}_J$ stands for $\mathbf{Q}_1\mathbf{A}_1$ and $\mathbf{Q}_J\mathbf{A}_J$ for $\mathbf{Q}_2\mathbf{A}_2$.Their solution to the problem is a transfer of normalization to $\mathbf{Q}_2\mathbf{A}_2$ when $\mathbf{Q}_1\mathbf{A}_1$ is in the process of being updated and vice versa, which implies that the orthogonality constraints are $\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = \mathbf{I}$ *or* $\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = \mathbf{I}$. Unfortunately, this cannot be applied in the present problem (we do not have to update $\mathbf{Y}_J$). Thus we have to attack the problem in a different way. Since the loss is minimized in for each set separately, the index $J$ will be omitted in the sequel.

In the first place we define a matrix $\mathbf{C}$ by $\mathbf{C} = (\mathbf{Q}'\mathbf{Q})^{1/2}\mathbf{A}$, so that $\mathbf{A} = (\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{C}$, with $\mathbf{A}'\mathbf{Q}'\mathbf{Q}\mathbf{A} = \mathbf{C}'\mathbf{C} = \mathbf{I}$. Now the metric projection problem (14) can be translated into

$$P_C(\mathbf{Y}) = \min_{\substack{\mathbf{q}_j \in C_j \\ \mathbf{C}'\mathbf{C}=\mathbf{I}}} \ \mathrm{tr}\,(\mathbf{Y} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{C})'(\mathbf{Y} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{C}), \tag{15}$$

In a first substep we minimize (15) over $\mathbf{C}$ for fixed $\mathbf{Q}$. This is an orthogonal Procrustes problem (Cliff, 1966), which is solved by defining the singular value decomposition

$$(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Y} = \mathbf{K}\mathbf{\Lambda}\mathbf{L}' \tag{16}$$

and by choosing $\tilde{\mathbf{C}}$ as $\mathbf{K}\mathbf{L}'$. In the next substep we have to transform the $\mathbf{h}_j$. In other normalized cone regression problems monotone regression is applied to columns of a matrix that has the same dimensionality as $\mathbf{Q}$. Part of the present problem is that $\mathbf{Y}$ is in general not of the dimensionality of $\mathbf{Q}$ (because we wish to find a reduced space). But the definition of $\mathbf{C}$ makes it possible to switch from minimizing (15) to the minimization of

$$P_C(\mathbf{Y}) = \min_{\substack{\mathbf{q}_j \in C_j \\ \mathbf{C}'\mathbf{C}=\mathbf{I}}} \ \mathrm{tr}\,(\mathbf{Y}\mathbf{C}' - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2})'(\mathbf{Y}\mathbf{C}' - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2}), \tag{17}$$

because

$$\begin{aligned} &\mathrm{tr}\,\mathbf{C}'(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{C} + \mathrm{tr}\,\mathbf{Y}'\mathbf{Y} - 2\,\mathrm{tr}\,\mathbf{C}'(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Y} = \\ &= r + \mathrm{tr}\,\mathbf{C}\mathbf{Y}'\mathbf{Y}\mathbf{C}' - 2\,\mathrm{tr}\,(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Y}\mathbf{C}', \end{aligned} \tag{18}$$

while

$$\begin{aligned} &\mathrm{tr}\,(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2} + \mathrm{tr}\,\mathbf{C}\mathbf{Y}'\mathbf{Y}\mathbf{C}' - 2\,\mathrm{tr}\,\mathbf{C}'(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Y} = \\ &= m + \mathrm{tr}\,\mathbf{C}\mathbf{Y}'\mathbf{Y}\mathbf{C}' - 2\,\mathrm{tr}\,(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{Q}'\mathbf{Y}\mathbf{C}', \end{aligned} \tag{19}$$

with $m$ the number of variables in the set and $r$ the chosen number of dimensions of the reduced space $\mathbf{QA}$. So (18) and (19) differ only by a constant. It is still not simple to minimize (17) because it is a normalized cone regression problem with a non-diagonal weight matrix that is dependent on the transformation. It is, however, the same problem as was mentioned in section 3. The problem can be solved by choosing $\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2}$ as the Gram-Schmidt transformation of $\mathbf{Q}$, taking care by permutation that the variable to be transformed is the last variable in the set, and by using majorization, i.e. replacement of the original function $\psi_1(\mathbf{x})$, which is not easily minimized, by a feasible quadratic function $\psi_2(\mathbf{x},\mathbf{y})$ for which it is true that for any $\mathbf{x}$ and $\mathbf{y}$ $\psi_1(\mathbf{x})\leq\psi_2(\mathbf{x},\mathbf{y})$ and $\psi_1(\mathbf{x})=\psi_2(\mathbf{x},\mathbf{y})=\psi_1(\mathbf{y})$ if $\mathbf{x}=\mathbf{y}$. When the optimal $\mathbf{Q}$ is obtained, we take $\mathbf{QA}$ as $\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1/2}\mathbf{C}$. When a new $\mathbf{Q}$ and $\mathbf{A}$ is obtained for each set, we return to (10) and compute a new $\mathbf{X}$. These steps are repeated until convergence is attained. When the variables are treated numerically, we only perform the steps (10), (12), (16).

## 5. APPLICATIONS

We return to the analysis of the radioactive fish. The ordinary generalized canonical analysis in section 2 has a proper stress of .865. Here it will be inspected how the technique described in the previous section behaves, whether we allow for transformation of the variables or not. Since we have to choose the dimensionality of the reduced canonical spaces, a principal components analysis was performed on the radioactivity measures set, and it turned out that there were 3 eigenvalues almost equal to zero, so the dimensionality of the reduced space was chosen to be 9-3=6. For set 2 the space was also reduced with three dimensions, which leaves a 4-dimensional space. The equivalent of multiple nominal treatment of the group variable in section 2 is taking the Mahalanobis distance based on the indicator matrix when minimizing a distance function (cf. Meulman, 1986). The indicator matrix $\mathbf{G}$ simply denotes by zeros and ones whether the fish belong to a particular aquarium. So the third set consists of the transformed variables $\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1/2}$, which is an $n$x3 matrix. In the first analysis the two other sets were treated numerically (proper stress .161), in the second analysis ordinally (proper stress .092). The two resulting configurations are shown in the Figures 3 and 4 respectively.

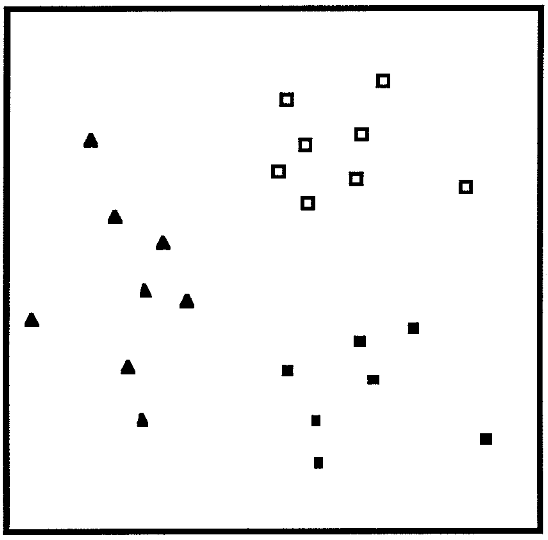


**Figure 3.** Two-dimensional configuration for radioactive fish ( ◻ shortest exposure, ▲ average exposure, ■ longest exposure). Numerical treatment of variables.

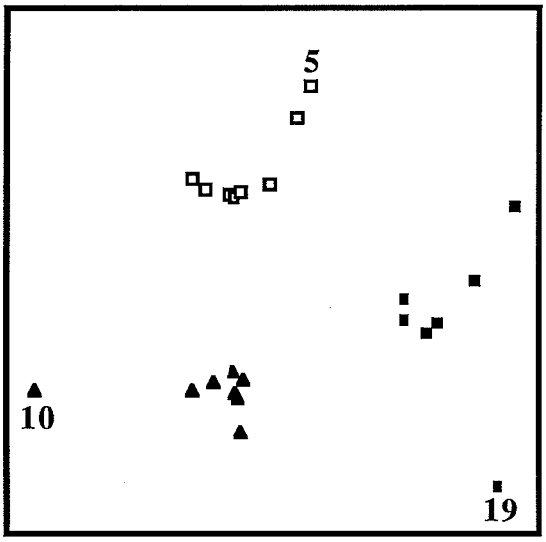



**Figure 4.** Two-dimensional configuration for radioactive fish ( ◻ shortest exposure, ▲ average exposure, ■ longest exposure). Ordinal treatment of variables.

Inspecting Figure 3 we see that the first dimension discrimates between the average exposed fish and the rest, while the second dimension separates the short and the long exposed (the configuration is in principal axes orientation, the eigenvalue of the first dimension is 1.743 and of the second 1.605). Comparing Figure 3 with Figure 1, it turns out that most of the distances (between *and* within) have become larger. The fit of the variables shows the same pattern as was described by the analysis in section 2: the average fit of the first set is .64, and of the second set .30, with exactly the same variables fitting very well and badly.

Before looking at Figure 4 we first inspect the plot of the residuals in the Figures 5 and 6. It is clear that the approximation is from both sides, since some fitted distances are larger than the distances to be approximated. This is especially true for the larger distances. In the Figure on the left the ■ indicate the distances that were badly fitted in the first analysis. In the Figure on the right we see that a lot of them are the within-group distances, which are also predominantly approximated from below by the reduced space analysis. In Figure 3 we see that there is quite a large distance between two fish in the average group. Very close inspection of the plot of the residuals shows that this distance is a very good approximation (there is one ■ in the middle of a lot of small dots just below the diagonal).

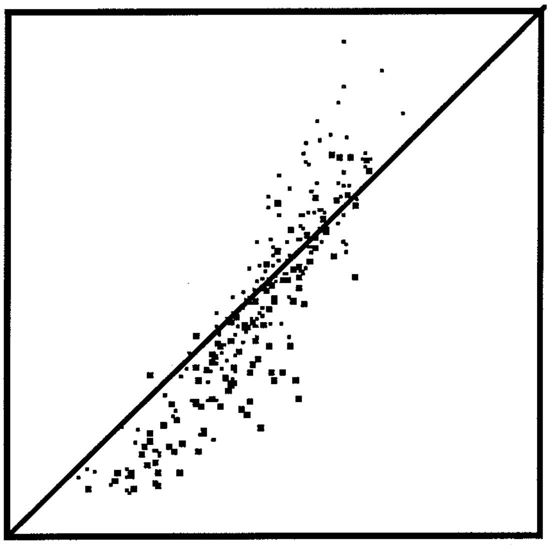


**Figure 5**. Plot of the residuals. Fitted distances versus average Mahalanobis distances in the reduced space analysis (linear). A ▪ indicates a distance that contributed more than average to the stress of the ordinary canonical analysis.

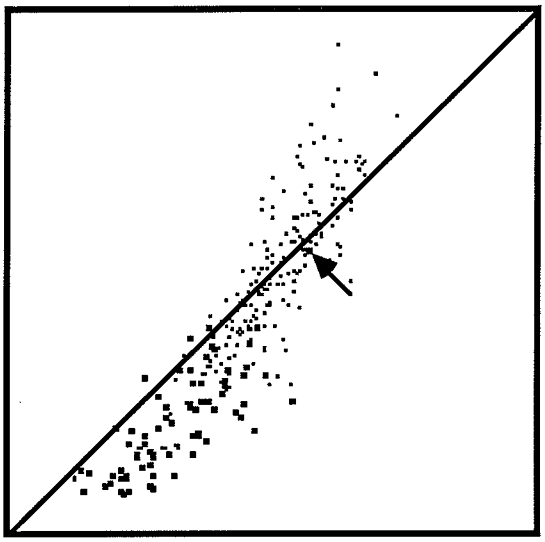


**Figure 6**. The same plot of the residuals, but now the ▪ indicate the within-group distances in the reduced space analysis.

Now we look at Figure 4, which gives the configuration of the reduced space analysis with ordinal transformations of the variables in the first two sets. We first remark that the transformed variables still fit in the space to the same agree as was seen in the two previous analyses (the average fit of the first set is .65, the five variables mentioned in section 2 also do fit very well here, but the mean fit is again very much influenced by the small fit of the variable that measures radioactivity of the kidneys (fit .080). The second set of variables has a average fit of .33. In the configuration the first dimension (eigenvalue 1.891) now separates the long exposed group from the rest, while the second dimension (eigenvalue 1.638) separates the short exposed from the rest: a very nice result. Most within-group distances have become very small, although there are a few remarkable exceptions. These fish already have large distances according to the original variables (and these were not always fitted very well before). But their extreme positions are predominantly caused by large between-group distances (and the ordinal transformations emphasize these aspects). We illustrate this in the Tables 1 up to 4, which give Mahalanobis and fitted distances between the fish 5, 10 and 19 and the members of their group, and between those three fish. The first two columns give figures for the ordinal reduced space analysis, the last two columns for the original canonical analysis.

**Table 1.** Column 1 and 3: Mahalanobis distances within a group in reduced space analysis and ordinary canonical analysis; column 2 and 4: corresponding fitted distances. All distances are between **fish 10** and its group members.

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| .68 | .51 | .63 | .19 |
| .64 | .51 | .53 | .09 |
| .64 | .50 | .72 | .26 |
| .63 | .49 | .55 | .17 |
| .63 | .49 | .69 | .17 |
| .58 | .44 | .55 | .10 |
| .54 | .38 | .65 | .29 |

**Table 2.** Identical to table 1, but here it concerns all distances between **fish 19** and its group members.

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| .83 | .69 | .61 | .25 |
| .69 | .52 | .60 | .17 |
| .61 | .51 | .70 | .23 |
| .62 | .47 | .64 | .17 |
| .61 | .43 | .52 | .04 |
| .62 | .41 | .50 | .09 |

**Table 3.** Identical to table 1, but here it concerns all distances between **fish 5** and its group members.

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| .55 | .36 | .60 | .22 |
| .51 | .36 | .47 | .18 |
| .46 | .33 | .45 | .17 |
| .44 | .33 | .46 | .11 |
| .43 | .31 | .70 | .11 |
| .35 | .26 | .39 | .16 |
| .12 | .09 | .29 | .13 |

**Table 4.** Identical to table 1, but here it concerns distances between fish that are **not** members of the same group.

| Fish | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 10-23 | .96 | 1.26 | .89 | .49 |
| 10-19 | 1.04 | 1.16 | .82 | .52 |
| 5-19 | .93 | 1.10 | .78 | .52 |
| 5-10 | .96 | 1.01 | .81 | .48 |

For the reduced space analysis with ordinal transformations we also give two residual plots, with the same labelling of points as in the Figures 5 and 6. Again it is true that most of the distances that are not well fitted by the reduced space analysis with linear transformations are the within-group distances, and they are still a little bit underestimated by the present analysis. The ordinal analysis is not complete without the transformation plots of the transformed variables versus the original variables. These are given in Figure 9a and 9b. In the radioactivity (RA) measures set the form of the transformations is not too irregular, except for RA of liver, RA of gullit and RA of kidneys (the variables that did not fit very well). We saw that the variables in the second set also did not fit very well, but their transformations do not illustrate this. bad fit.

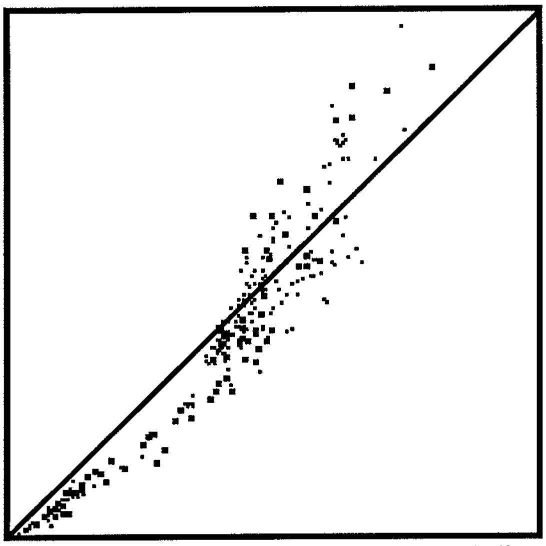


**Figure 7**. Plot of the residuals. Fitted distances versus average Mahalanobis distances in the reduced space analysis (ordinal). A ▪ indicates a distance that contributed more than average to the stress of the linear reduced space analysis.

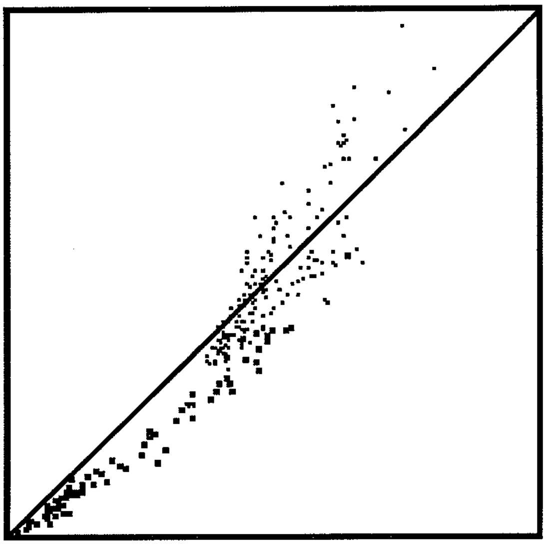
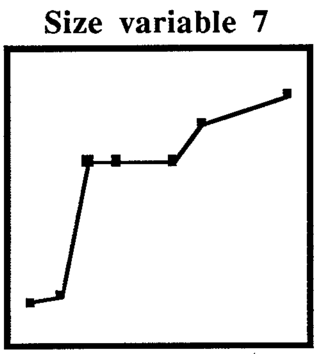


**Figure 8**. The same plot of the residuals, but now the ▪ indicate the within-group distances in the nonlinear reduced space analysis.




Interpretation of a transformation plot. The original values are along the horizontal axes, and the transformed values along the vertical axis. So size variable 7 had originally 7 different values; in the transformation the third, fourth and fifth value were made equal.

**Figure 9a**. Transformation of variable 7 of set 2.

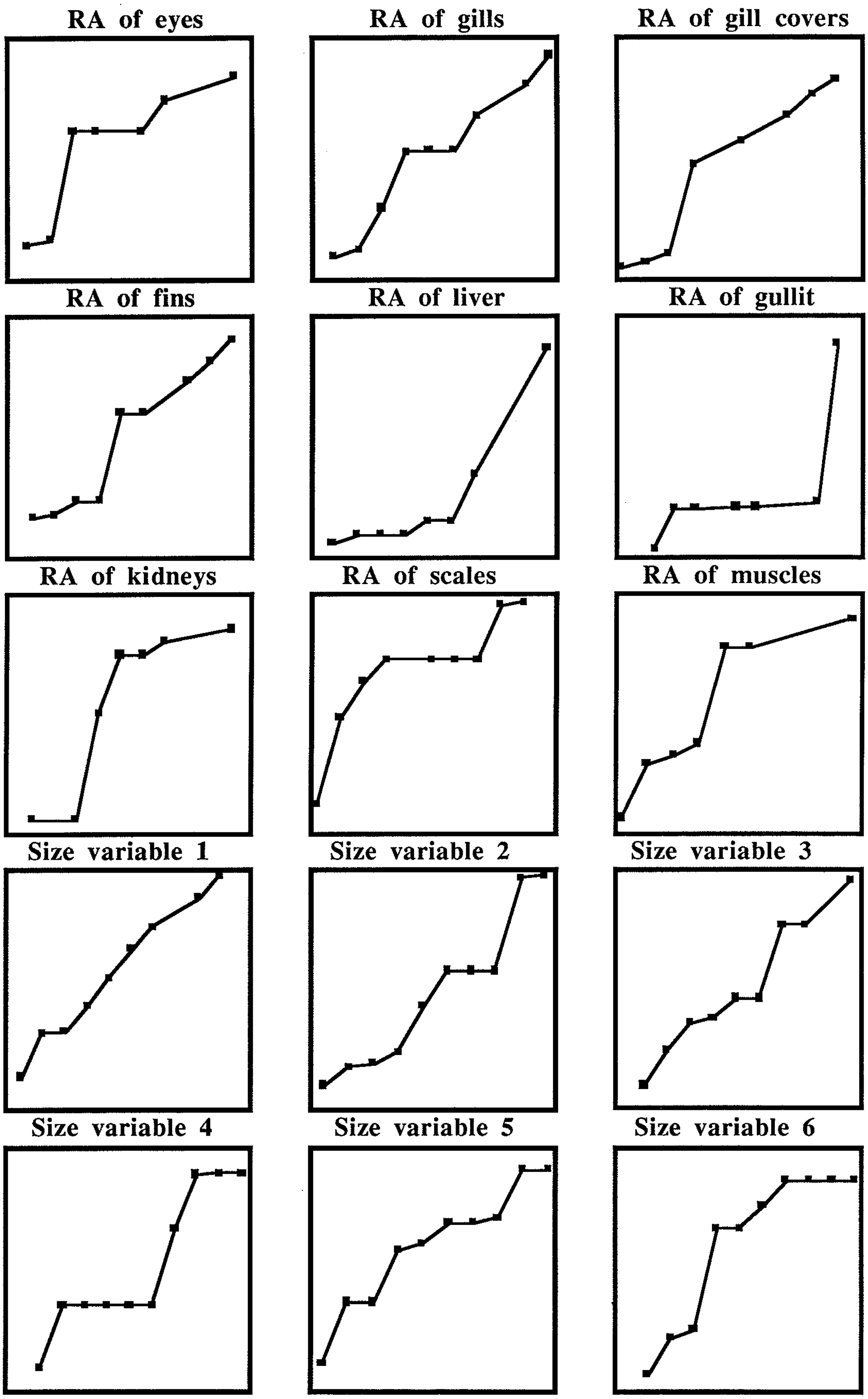



**Figure 9b.** Transformations of the 9 variables of set 1 and 6 variables of set 2.

## 6. DISCUSSION

We summarize some of the reasons why least squares distance fitting in reduced canonical spaces is a candidate substitute of the more traditional approaches to generalized canonical analysis. If it is especially the objects that we are interested in, then we obtain a somewhat more fair representation of the objects in terms of distances. Moreover,we acquire a lot of flexibility in the choice of the dimensionality of the canonical spaces. In the first place we could choose the dimensionality in between $p$, the dimensionality of the comparison space, and $m_J$, the dimensionality of set $J$. In this case one would hope that the reduced space contains the more important information, while unimportant information has been neglected. In the second place, we could choose the dimensionality to be smaller than $p$, for instance equal to 1. This could be desirable in the following three-way situation. Suppose we have an $nxmxt$ data matrix as shown in Figure 10. The $n$ rows indicate a set of objects, the $m$ columns a group of judges, and the third way gives judgements on $t$ different characteristics. We would like to have an aggregated opinion across the judges with respect to a particular characteristic (a single dimension). But the structure of the characteristics themselves need not to be one-dimensional. This calls for a more-dimensional comparison space **X**.

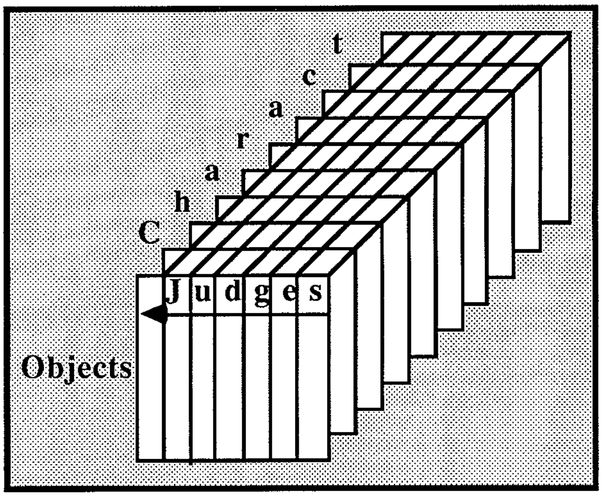



**Figure 10.** Multiway situation in which reduced space analysis is feasible.

Finally it should be noted that the technique can be extended in a number of ways. The different sets could be given a priori differential weights, for instance dependent on the number of canonical variables that are retained. Another very interesting extension is motivated by the intrinsic relationship between individual differences scaling via Carroll and Chang's (1970) CANDECOMP method and generalized canonical analysis (a formal description of this relationship was given in Heiser and Meulman, 1983). Incorporating individual differences models would directly relate the axes of the comparison space with the canonical spaces, giving us weights for each set. To attain this objective the distance approach to canonical analysis described in this paper should be integrated with the SMACOF approach to individual differences modeling as described in Heiser and Stoop (1986).

## REFERENCES

Cailliez, F. and Pagès, J.-P. (1976). *Introduction à l'analyse des données*. Paris: SMASH.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th annual convention of the American Psychological Association, 3*, 227-228.

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*, 283-391.

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

Dauxois, J. and Pousse, A. (1976). *Les analyses factorielles en calcul des probabilités et en statistique: essai d'étude synthétique* [Thèse de Doct. 3e cycle]. Toulouse: Université Paul Sabatier.

De Leeuw, J. (1984). The Gifi system of nonlinear multivariate analysis. In: E. Diday et al. (Eds.), *Data analysis and informatics, III*, 415-424. Amsterdam: North-Holland.

Gifi, A. (1981). *Nonlinear multivariate analysis*. Leiden: Dept. of Data Theory. A revised edition will be published in 1988 by DSWO Press, Leiden.

Gower, J.C. (1966). Some distance properties of latent roots and vector methods used in multivariate analysis. *Biometrika, 53*, 325-388.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika, 33*, 469-506.

Heiser, W. J. and De Leeuw, J. (1979). How to use SMACOF-III. *Research Report*, Leiden: Dept. of Data Theory.

Heiser, W. J. and Meulman, J. (1983). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement, 7*, 381-404.

Heiser, W. J. and Stoop, I. (1986). Explicit SMACOF algorithms for individual differences scaling. *Research Report RR-86-14*. Leiden: Dept. of Data Theory.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika, 58*, 433-460.

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

Kruskal, J. B. Young, F. W. and Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding*. Murray Hill, NJ: Bell Laboratories.

Meulman, J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.

Meulman, J. (1988). Reduced space analysis. *Research Report RR-88-01*. Leiden: Dept. of Data Theory.

Ramsay, J. O. (1982b). *MULTISCALE II Manual*. McGill University, Montréal: Dept. of Psychology.

Saporta, G. (1975). *Liaisons entre plusieurs ensembles de variables et codage de données qualitatives* [Thèse de Doct. 3e cycle]. Paris: Université Paris VI.

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

Van de Geer, J. P. (1984). Relations among K sets of variables. *Psychometrika, 49*, 79-94.

Van de Geer, J. P. (1986). *Introduction to linear multivariate data analysis* (2 vols). Leiden: DSWO Press.

Van der Burg, E. and De Leeuw, J. (1983). Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology, 36*, 54-80.

Van der Burg, E., and De Leeuw, J. (1987). Nonlinear canonical correlation with *k* sets of variables. *Research Report RR-87-8*. University of Twente: Dept. of Education.

Van der Burg, E., De Leeuw, J. and Verdegaal, R. (1984). Nonlinear canonical correlation with *M* sets of variables. *Research Report RR-84-12*. Leiden: Dept. of Data Theory.

Van der Burg, E., De Leeuw, J. and Verdegaal, R. (1986). Homogeneity analysis with $K$ sets of variables: an alternating least squares method with optimal scaling features. Research Report RR-86-5. University of Twente: Department of Education. To appear in *Psychometrika*.

Van der Burg, E., and Dijksterhuis, G. (1988). Nonlinear canonical correlation analysis of multiway data. [In these proceedings].

Young, G. and Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, *3*, 19-22.