Generalized canonical correlation analysis of matrices with different row and column orders

Michel van de Velden*and Tammo H.A. Bijmolt†

June 2003

Abstract

A method is offered that makes it possible to apply generalized canonical correlations analysis (CANCOR) to two or more matrices of different row and column order. The new method optimizes the generalized canonical correlation analysis objective by considering only the observed values. This is achieved by employing selection matrices. We present and discuss fit measures to assess the quality of the solutions. In a simulation study we assess the performance of our new method and compare it to an existing procedure called GENCOM, proposed by Green and Carroll. We find that our new method outperforms the GENCOM algorithm both with respect to model fit and recovery of the true structure. Moreover, as our new method does not require any type of iteration it is easier to implement and requires less computation. We illustrate the method by means of an example concerning the relative positions of political parties in the Netherlands based on provincial data.

Key words: Generalized canonical correlation analysis, perceptual mapping.

JEL classification: C19, C88, M31.

*Departament d'Economia i Empresa, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain; phone: +34 935421942; fax: +34 935421746; email: michel.vandevelden@upf.edu. The research of Michel van de Velden was supported by EU grant HPMF-CT-2000-00664

†Department of Marketing, Tilburg University, PO box 90153, 5000 LE, Tilburg, The Netherlands; phone: +31 134663423; fax: +31 134662875; email: t.h.a.bijmolt@uvt.nl

1 Introduction

Carroll (1968) introduced a generalization of canonical correlation analysis that allows the simultaneous analysis of n sets of variables. His method is referred to as CANCOR. In CANCOR linear combinations of the variables of the individual observation matrices are chosen in such a way that the sum of the squared correlations between the linear combinations and an (unknown) group configuration is at a maximum. Some important advantages of CANCOR are its ease of computation and the flexibility of the input data. In particular, the fact that the columns of the individual observation matrices may correspond to different variables and that the number of columns for each observation matrix may differ, makes the method widely applicable. Steenkamp, Van Trijp, and Ten Berge (1994), for example, make use of this flexibility in their analysis of data in which individuals evaluated a set of objects using only criteria that they deem important. Another possible application concerns the simultaneous analysis of several multidimensional scaling solutions to derive a common configuration. The multidimensional scaling solutions (i.e. the coordinate matrices) may then again be of different dimensionality. Hence, the CANCOR model can be applied in a range of research settings.

In this paper we generalize the CANCOR method such that it can be applied to data where not only the number of columns for each individual data matrix may differ, but also the number of rows. Thus, individual observation matrices can be of different row and column orders. This type of data is particularly interesting as it allows, for example, the simultaneous analysis of several multidimensional scaling solutions obtained at different time points (or from different regions) where not all objects are observed at all times (or in all regions). Or, we may consider data where individuals evaluate only objects that they deem relevant (or that they are familiar with), using only the attributes that they find relevant.

The new method that we propose does not require any type of iteration and yields immediate solutions based on the observed values alone. In that sense, it differs from a method proposed by Green and Carroll (1988) to deal with missing observations in CANCOR. In the Green and Carroll approach missing values are estimated and a solution is obtained using an iterative algorithm which they call GENCOM.

The performance of our new method will be assessed using a simulation study. By varying several factors, such as the number of missing rows and columns, and the number of observations, we evaluate the performance of our method under various conditions. Also, we will use the simulation study to compare our new method with the existing GENCOM algorithm. Finally, the methods are illustrated using an example on the positioning of political parties in the Netherlands.

The paper is organized as follows. First we briefly introduce CANCOR using a similar formulation as employed by Steenkamp, Van Trijp, and Ten Berge (1994). Then, in Section 3, we introduce our new approach for generalized canonical correlation analysis applied to data with different row and column orders. Measures to assess the quality of the solutions are given in Section 4. In Section 5 we briefly present Green and Carroll's (1988) algorithm and in Section 6 the simulation study and its results are described. An application concerning the political landscape in

the Netherlands is given in Section 7 and the paper is concluded with a discussion of our results.

2 Generalized canonical correlation analysis: CANCOR

Suppose we have n column centered data matrices $\mathbf{X}_i$ of orders $m \times p_i$ where the rows for the different data matrices correspond to the same objects. The $\mathbf{X}_i$ may for example correspond to coordinate matrices obtained using multidimensional scaling methods, or they may represent (centered) observation matrices where (the same) m objects are evaluated using p_i attributes. We can analyze such data using CANCOR, a generalization of canonical correlation analysis introduced by Carroll (1968). In CANCOR, we obtain an orthogonal k dimensional group configuration $\mathbf{Y}$ and k linear combinations of the centered data matrices $\mathbf{X}_i$ in such a way that the sum of the squared distances between the group configuration $\mathbf{Y}$ and the linear combinations is minimized.

The objective of CANCOR can be expressed in the following way:

$$\min_{\mathbf{A},\mathbf{Y}} \ \phi = \text{trace} \sum_{i=1}^{n} \left(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i\right)' \left(\mathbf{Y} - \mathbf{X}_i\mathbf{A}_i\right) \tag{1}$$

$$s.t.\ \mathbf{Y}'\mathbf{Y} = \mathbf{I}_k. \tag{2}$$

It is known, e.g. Carroll (1968), that the group configuration matrix $\mathbf{Y}$ can be obtained using the eigenequation

$$\left(\sum_{i=1}^{n} \mathbf{X}_i \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\right) \mathbf{Y} = \mathbf{Y}\mathbf{\Lambda}, \tag{3}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements λ_j, being the k largest eigenvalues of $\sum_{i=1}^{n} \mathbf{X}_i \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'$ (where we have assumed that the $\mathbf{X}_i's$ are of full column rank) and the columns of $\mathbf{Y}$ are corresponding eigenvectors. The matrices $\mathbf{A}_i$ can be calculated as

$$\mathbf{A}_i = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\mathbf{Y}. \tag{4}$$

The computational ease as well as the flexibility in the dimensionality of the input matrices $\mathbf{X}_i$ make the method very appealing for several types of applications. Green and Carroll (1988), for example, apply the method to obtain a composite of several multidimensional scaling solutions. Steenkamp, Van Trijp, and Ten Berge (1994) use the method to obtain a common perceptual map based on consumer's judgments of brands using idiosyncratic sets of attributes.

Squared correlations The eigenvalues in $\mathbf{\Lambda}$ are equal to the average squared correlation between the dimensions of the group configuration $\mathbf{Y}$ and the linear combinations $\mathbf{X}_i\mathbf{A}_i$. That is,

$$\lambda_j = \frac{1}{n} \sum_{i=1}^{n} \rho_{ij}^2,$$

where ρ_{ij}^2 denotes, for the ith observation matrix, the squared correlation between the jth linear combination of $\mathbf{X}_i$ and the jth dimension of $\mathbf{Y}$. Hence,

$$\rho_{ij}^2 = \frac{\left(\mathbf{y}_j'\mathbf{X}_i\mathbf{a}_{i(j)}\right)^2}{\left(\mathbf{y}_j'\mathbf{y}_j\right)\left(\mathbf{a}_{i(j)}'\mathbf{X}_i'\mathbf{X}_i\mathbf{a}_{i(j)}\right)} = \mathbf{y}_j'\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{y}_j,$$

where $\mathbf{y}_j$ denotes the jth column of $\mathbf{Y}$, $\mathbf{a}_{i(j)}$ is the jth column of $\mathbf{A}_i$ and we used (2) and (4). It should be noted that Carroll (1968) formulates CANCOR as the optimization of the sum (over i) of these squared correlations.

Another way to interpret the squared correlations is obtained by observing the relationship with regression analysis. For the ith observation matrix and the jth dimension we can write

$$\mathbf{y}_j = \mathbf{X}_i\mathbf{a}_{i(j)} + \mathbf{e}_{ij}, \tag{5}$$

where $\mathbf{e}_{ij}$ is a $k \times 1$ vector of residuals. Hence, the elements of $\mathbf{A}_i$ can be interpreted as regression coefficients from the ordinary least squares regression of $\mathbf{y}_j$ on $\mathbf{X}_i$. Using this relationship it is not difficult to see that the squared correlation ρ_{ij}^2 is equivalent to the multiple correlation coefficient R^2 obtained in the regression analysis (5).

3 Generalized Canonical Correlation Analysis with unequal number of rows

As shown in the previous section, we can apply CANCOR to obtain a group configuration for a set of m objects based on n individual observation matrices. Furthermore, the number of attributes used to evaluate the objects may differ per observation matrix (i.e. the number of columns of $\mathbf{X}_i$ is individual specific). The number of objects, on the other hand, must always be equal to the size of the complete set. This implies that, if we are, for example, comparing several multidimensional scaling configurations, each configuration consists of the same set of objects.

The situation, however, may occur that for each observation matrix $\mathbf{X}_i$ only data on a subset of the m objects is available. Hence, instead of n matrices of order $m \times p_i$ we have n matrices of orders $m_i \times p_i$. That is, the row and column orders of the $\mathbf{X}_i$ matrices are individual specific. In this section we will extent the CANCOR approach in such a way that a group configuration for the set of m objects can be obtained based on the $m_i \times p_i$ observation matrices.

Like before, we are interested in obtaining a group configuration that summarizes the complex relationships between the objects based on the observations matrices of different dimensionalities. Therefore, we derive the group configuration $\mathbf{Y}$ in such a way that the sum (over the n subjects) of the distances between the coordinates for the *observed* objects and the corresponding linear combinations of the attributes is minimized. Thus, we will only consider the distances (differences) between the linear combinations of the observed m_i rows of $\mathbf{X}_i$ and the corresponding rows in the $m \times k$ group configuration $\mathbf{Y}$. This can be achieved by employing selection matrices.

If we have m_i observed rows for individual i we introduce an $m \times m$ selection matrix $\mathbf{K}_i$ with elements

$$K_{jj}^i = \begin{cases} 1 & \text{if we have an observation for individual } i \text{ in the } j\text{th row} \\ 0 & \text{else} \end{cases}.$$

It is easily verified that these selection matrices are idempotent, i.e. $\mathbf{K}_i\mathbf{K}_i = \mathbf{K}_i$. By adding $m - m_i$ rows to the observed matrices $\mathbf{X}_i$ (in such a way that the rows of the different $\mathbf{X}_i$'s correspond) we obtain n matrices, say $\mathbf{X}_i^*$, of order $m \times p_i$. Premultiplying $\mathbf{Y}$ and $\mathbf{X}_i^*$ with $\mathbf{K}_i$ yields matrices with zero-rows corresponding to the unobserved objects.

Using the idempotent selection matrices $\mathbf{K}_i$, we can formulate as objective

$$\min_{\mathbf{A}_i, \mathbf{Y}} \ \phi = \text{trace} \sum_{i=1}^{n} \left(\mathbf{K}_i\mathbf{Y} - \mathbf{K}_i\mathbf{X}_i^*\mathbf{A}_i\right)' \left(\mathbf{K}_i\mathbf{Y} - \mathbf{K}_i\mathbf{X}_i^*\mathbf{A}_i\right), \tag{6}$$

and we will minimize this objective subject to the constraint $\mathbf{Y}'\mathbf{K}\mathbf{Y} = n\mathbf{I}_k$, where $\mathbf{K} = \sum_{i=1}^{n} \mathbf{K}_i$.

The standardization $\mathbf{Y}'\mathbf{K}\mathbf{Y} = n\mathbf{I}_k$ implies that objects which are relatively infrequently observed receive larger weights than objects that are relatively often observed. This weighting of the objects is similar to the weighting of coordinates in, for example, correspondence analysis (see, e.g., Greenacre, 1984). It prevents the configuration from being dominated by objects that are relatively often observed. Moreover, if each subject observes all m objects, that is, $\mathbf{K}_i = \mathbf{I}_m$, for $i = 1, \ldots, n$, our approach is identical to CANCOR.

Assuming, for the moment, that all matrices $\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*$ are non-singular, we find that the group configuration can be obtained from the following eigenequation,

$$\mathbf{K}^{-\frac{1}{2}}\mathbf{X}\mathbf{D}^{-1}\mathbf{X}'\mathbf{K}^{-\frac{1}{2}}\mathbf{Y}^* = \mathbf{Y}^*\mathbf{\Lambda}, \tag{7}$$

where $\mathbf{X} = \begin{pmatrix} \mathbf{K}_1\mathbf{X}_1^* & \mathbf{K}_2\mathbf{X}_2^* & \cdots & \mathbf{K}_n\mathbf{X}_n^* \end{pmatrix}$

$\mathbf{D} = \begin{pmatrix} \mathbf{X}_1^{*\prime}\mathbf{K}_1\mathbf{X}_1^* & & & \\ & \mathbf{X}_2^{*\prime}\mathbf{K}_2\mathbf{X}_2^* & & \\ & & \ddots & \\ & & & \mathbf{X}_n^{*\prime}\mathbf{K}_n\mathbf{X}_n^* \end{pmatrix}$, $\mathbf{Y}^* = \frac{1}{\sqrt{n}}\mathbf{K}^{\frac{1}{2}}\mathbf{Y}$ and the objective is minimized when the largest eigenvalues and corresponding eigenvectors are selected. (A complete derivation using matrix derivatives can be found in appendix A.)

The individual linear combinations can be calculated using

$$\mathbf{A} = \mathbf{D}^{-1}\mathbf{X}'\mathbf{Y}, \tag{8}$$

where $\mathbf{A} = \begin{pmatrix} \mathbf{A}_1' & \mathbf{A}_2' & \cdots & \mathbf{A}_n' \end{pmatrix}'$.

Alternatively, we can write, in accordance with (3) and (4),

$$\mathbf{K}^{-\frac{1}{2}} \left(\sum_{i=1}^{n} \mathbf{K}_i\mathbf{X}_i^* \left(\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*\right)^{-1} \mathbf{X}_i^{*\prime}\mathbf{K}_i \right) \mathbf{K}^{-\frac{1}{2}}\mathbf{Y}^* = \mathbf{Y}^*\mathbf{\Lambda},$$

and

$$\mathbf{A}_i = \left(\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*\right)^{-1} \mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{Y}.$$

Rank problems Thus far we have assumed that the matrices $\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*$ are non-singular. This, however, need not be the case. If a subject for example evaluates m_i objects using p_i attributes with $m_i < p_i$ then the rank of $\mathbf{K}_i\mathbf{X}_i^*$ is smaller than or equal to m_i. Hence, $\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*$ is singular. Although a similar problem can occur in CANCOR, i.e. when the rank of $\mathbf{X}_i$ is smaller than p_i, it receives no attention in the literature. A reason for this is perhaps the fact that in the CANCOR case the problem may be avoided by taking the number of rows, i.e. m, sufficiently large. Here, however, we do not wish to impose any explicit restrictions on the $m_i's$.

To resolve the singularity problem we consider the singular value decomposition of $\mathbf{K}_i\mathbf{X}_i^*$, i.e.

$$\mathbf{K}_i\mathbf{X}_i^* = \mathbf{U}_i\mathbf{\Phi}_i\mathbf{V}_i', \tag{9}$$

where $\mathbf{U}_i'\mathbf{U}_i = \mathbf{V}_i'\mathbf{V}_i = \mathbf{I}_{\kappa_i}$ and κ_i denotes the rank of $\mathbf{K}_i\mathbf{X}_i^*$. Now, instead of using the $m \times p_i$ data matrix $\mathbf{K}_i\mathbf{X}_i^*$, we may use the linear combinations obtained by postmultiplying $\mathbf{K}_i\mathbf{X}_i^*$ with $\mathbf{V}_i$, i.e.

$$\breve{\mathbf{X}}_i = \mathbf{K}_i\mathbf{X}_i^*\mathbf{V}_i, \tag{10}$$

which is an $m \times \kappa_i$ matrix of full column rank. Thus, the dependencies among the columns of the original data matrix are used to obtain a matrix that has full column rank.

By substituting $\breve{\mathbf{X}}_i$ for $\mathbf{K}_i\mathbf{X}_i^*$ we have decreased the dimensionality of the original data matrix in such a way that the singularity problems no longer occur and the calculation of the group configuration becomes straightforward. However, using $\breve{\mathbf{X}}_i$ we obtain linear combinations for the κ_i columns of $\breve{\mathbf{X}}_i$ rather than for the original p_i columns of $\mathbf{K}_i\mathbf{X}_i^*$. For the calculation of the group configuration matrix $\mathbf{Y}$ this does not matter. However, if we are interested in the linear combinations for the original p_i columns, some additional steps are required.

Let $\tilde{\mathbf{A}}_i$ denote the matrix of linear combinations based on the $\breve{\mathbf{X}}_i$ matrices, i.e.,

$$\begin{aligned}
\tilde{\mathbf{A}}_i &= \left(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i\right)^{-1}\breve{\mathbf{X}}_i'\mathbf{Y} \\
&= \left(\mathbf{V}_i'\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*\mathbf{V}_i\right)^{-1}\mathbf{V}_i'\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{Y} \\
&= \mathbf{\Phi}^{-1}\mathbf{U}_i'\mathbf{Y},
\end{aligned} \tag{11}$$

where we used (9), (10), and the idempotency of $\mathbf{K}_i$. Now, in order to obtain linear combinations for the p_i columns of $\mathbf{K}_i\mathbf{X}_i^*$ we need to pre-multiply $\tilde{\mathbf{A}}_i$ by $\mathbf{V}_i$. Using (9) and (11) we obtain

$$\mathbf{A}_i = \mathbf{V}_i\tilde{\mathbf{A}}_i = \mathbf{V}_i\mathbf{\Phi}^{-1}\mathbf{U}_i'\mathbf{Y} = \mathbf{V}_i\mathbf{\Phi}_i^{-2}\mathbf{V}_i'\mathbf{V}_i\mathbf{\Phi}_i\mathbf{U}_i'\mathbf{K}_i\mathbf{Y} = \left(\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*\right)^{+}\mathbf{X}_i^{*'}\mathbf{K}_i\mathbf{Y}. \tag{12}$$

Because,

$$\left(\mathbf{X}_i^{*\prime}\mathbf{K}_i\mathbf{X}_i^*\right)^{+} = \mathbf{V}_i\mathbf{\Phi}_i^{-2}\mathbf{V}_i',$$

where, generically, $\mathbf{B}^{+}$ denotes the Moore Penrose inverse of $\mathbf{B}$. Thus, instead of taking the linear combinations (10), we may simply replace the inverses in (7) and (8) by the Moore Penrose-inverses.

4 Assessing the quality of the CANCOR solution

4.1 Squared correlations

For each subject we can calculate squared correlations between the linear combinations of the observations and the jth dimension of the group configuration as

$$\rho_{ij}^2\left(\mathbf{K}_i\mathbf{y}_j, \mathbf{K}_i\mathbf{X}_i\mathbf{a}_{(i)j}\right) = \frac{\left(\mathbf{y}_j'\mathbf{K}_i\mathbf{X}_i\mathbf{a}_{i(j)}\right)^2}{\left(\mathbf{y}_j'\mathbf{K}_i\mathbf{y}_j\right)\left(\mathbf{a}_{i(j)}'\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i\mathbf{a}_{i(j)}\mathbf{y}_j\right)} = \frac{\mathbf{y}_j'\mathbf{K}_i\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{K}_i\mathbf{y}_j}{\mathbf{y}_j'\mathbf{K}_i\mathbf{y}_j},$$

where, $\mathbf{y}_j$ denotes the jth column of $\mathbf{Y}$, $\mathbf{a}_{i(j)}$ is the jth column of $\mathbf{A}_i$ and, for convenience, we have dropped the superscribed $*'s$. If $\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i$ is singular we replace the regular inverse by the Moore Penrose inverse. Note that $\mathbf{y}_j'\mathbf{K}_i\mathbf{y}_j$ is not necessarily equal to one. Consequently, the sum, over i, of the squared correlations between the linear combinations $\mathbf{X}_i\mathbf{a}_{i(j)}$ and the jth dimension of the group configuration $\mathbf{Y}$, is *not* necessarily equal to the jth diagonal element of $\mathbf{\Lambda}$. Instead, the eigenvalues λ_j are equal to weighted sums of squared correlations where the weights are $\left(\mathbf{y}_j'\mathbf{K}_i\mathbf{y}_j\right)^{-1}$ for $i = 1, \ldots n$. These weighted sums of squared correlations are maximized.

Another way to interpret the squared correlations is obtained by observing the analogy with regression analysis. For the ith individual and the jth dimension we have

$$\mathbf{K}_i\mathbf{y}_j = \mathbf{K}_i\mathbf{X}_i\mathbf{a}_{i(j)} + \mathbf{e}_{ij}, \tag{13}$$

where $\mathbf{e}_{ij}$ is a vector of residuals. Hence, the elements of $\mathbf{A}_i$ can be interpreted as regression coefficients, and the squared correlation $\rho_{ij}^2\left(\mathbf{K}_i\mathbf{y}_j, \mathbf{K}_i\mathbf{X}_i\mathbf{a}_{i(j)}\right)$ is equivalent to the multiple correlation coefficient R^2 obtained in the ordinary least squares regression of $\mathbf{K}_i\mathbf{y}_j$ and $\mathbf{K}_i\mathbf{X}_i$.

4.2 Redundancy

In order to assess the quality of the solution it would also be interesting to see how well the group configuration represents the individual data. For this purpose we will use the redundancy index introduced by Stewart and Love (1968). The redundancy index can be seen as a multivariate extension of the multiple correlation coefficient. It is defined as the variance of the predicted variables divided by the variance of the predictor variables. Here we will use the redundancy index to assess how well the original data are represented by the k dimensional group configuration. That is, we will, for each individual, consider the prediction of $\mathbf{K}_i\mathbf{X}_i$ by $\mathbf{K}_i\mathbf{Y}$.

For the prediction of $\mathbf{K}_i\mathbf{X}_i$ by $\mathbf{K}_i\mathbf{Y}$, we consider the following model: $\mathbf{K}_i\mathbf{X}_i = \mathbf{K}_i\mathbf{Y}\mathbf{B}_i + \mathbf{E}$, where $\mathbf{B}_i$ is a matrix of coefficients and $\mathbf{E}$ is a matrix of residuals. Hence, $\mathbf{B}_i = \left(\mathbf{Y}'\mathbf{K}_i\mathbf{Y}\right)^{-1}\mathbf{Y}'\mathbf{K}_i\mathbf{X}_i$ and $\hat{\mathbf{X}}_i = \mathbf{K}_i\mathbf{Y}\mathbf{B}_i$. The redundancy index for the ith individual, defined as the variance of the predicted variables divided by the variance of the predictor variables, can be calculate as

$$RI_{X_i|Y} = \frac{Var(\hat{\mathbf{X}}_i)}{Var\left(\mathbf{K}_i\mathbf{X}_i\right)} = \frac{\text{trace}\left(\mathbf{X}_i'\mathbf{K}_i\mathbf{Y}\left(\mathbf{Y}'\mathbf{K}_i\mathbf{Y}\right)^{-1}\mathbf{Y}'\mathbf{K}_i\mathbf{X}_i\right)}{\text{trace}(\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i)}.$$

A high redundancy index indicates that a large amount of variation present in the individual's original data matrix $\mathbf{X}_i$ is accounted for by the group configuration $\mathbf{Y}$. An overall measure of fit can be obtained by calculating the average redundancy index

$$\overline{RI}_{X|Y} = \frac{1}{n}\sum_{i=1}^{n} RI_{X_i|Y}. \tag{14}$$

Steenkamp, Van Trijp, and Ten Berge (1994) use a similar measure of fit which they call variance accounted for (VAF). They calculate the VAF as the average of the multiple correlation coefficients obtained in the regressions of $\mathbf{x}_{i(j)}$, where $\mathbf{x}_{i(j)}$ denotes the jth column of $\mathbf{X}_i$, on $\mathbf{Y}$. It is not difficult to see that for standardized variables, that is if $\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i = \mathbf{I}_{p_i}$ the average redundancy index and VAF are the same.

4.3 Determining the dimensionality of the group configuration

The choice of the dimensionality of the group configuration depends on the interpretability of the solution and on the amount of variation accounted for. In particular, the number of dimensions must be such that a substantial part of the variation in the data is described. There does not appear to exists a clear cut procedure for determining the dimensionality of the group configuration $\mathbf{Y}$, however, we can use the various measures described in the previous sections to select the number of dimensions that are needed.

As described in Section 4.1, λ_k, i.e. the kth diagonal element of $\mathbf{\Lambda}$, is a weighted sum of squared correlations between the linear combinations $\mathbf{X}_i\mathbf{a}_{i(k)}$ and the kth dimension of the group configuration. Relatively low values for λ_k indicate that, on average, the kth linear combinations are not highly correlated with the kth dimension of the group configuration. Thus, the kth and, as the diagonal elements of $\mathbf{\Lambda}$ are non-increasing, subsequent dimensions, can be considered superfluous. To determine what is "relatively low" one may, as is often done in for example principal component analysis, plot the λ's and look for an appropriate number of dimensions.

An alternative measure that one can consider is the average redundancy index given in (14). From the close relationship between the multiple correlation coefficient and the redundancy index it follows that adding dimensions to the group configuration will lead to an increase in the redundancy indices $RI_{X_i|Y}$ and hence in $\overline{RI}_{X|Y}$. If dimensions that describe little variance are added, the increase in the average redundancy index will be small. Therefore, in a similar way as described for the eigenvalues λ_k, we can consider the average redundancy index $\overline{RI}_{X|Y}$ for solutions of different dimensionality and determine the number of dimensions. Note however, that, unlike the eigenvalues λ_k, the redundancy indices for different dimensionality must be calculated separately. That is, for different k values the redundancy indices must be calculated as described in the previous section.

5 Green and Carroll's GENCOM algorithm

Green and Carroll (1988) proposed an iterative procedure for dealing with missing elements in CANCOR. Their method is more general than the approach presented here in the sense that it can also be used when only certain cells, rather than complete rows, of the $\mathbf{X}_i^*$ matrices are missing. The basic principle in their approach, which they call GENCOM, is to treat the non-observed rows as missing values and to estimate these missing values using an iterative procedure. Green and Carroll (1988) do not give details on numerical properties of the algorithm.

When complete rows are missing the GENCOM algorithm can be formulated as follows:

1. For each $\mathbf{X}_i^*$ calculate $\hat{\mathbf{X}}_i^{(t)}$ by replacing the missing values by the column averages. Thus, the elements of the rows of $\hat{\mathbf{X}}_i^{(t)}$ that correspond to missing values, become: $\frac{1}{m_i}\mathbf{X}_i^{*\prime}\mathbf{1}_m$. (The superscribed t's correspond to iterations).

2. Construct $\mathbf{Y}^{(t)}$ by applying CANCOR to the $\hat{\mathbf{X}}_i^{(t)}$ matrices and by supplementing a column of ones (for estimating the intercept term) to the thus obtained configuration matrix.

3. For each $\mathbf{X}_i$ use ordinary least-squares to fit: $\mathbf{X}_i = \mathbf{Y}^{(t)*}\mathbf{B}_i^{(t)}$, where $\mathbf{X}_i$ is the original data matrix of order $m_i \times p_i$, and $\mathbf{Y}^{(t)*}$ is the matrix of corresponding rows of $\mathbf{Y}^{(t)}$. Thus, $\mathbf{B}_i^{(t)} = \left(\mathbf{Y}^{(t)*\prime}\mathbf{Y}^{(t)*}\right)^{-1}\mathbf{Y}^{(t)*\prime}\mathbf{X}_i$.

4. Let $\mathbf{X}_i^{(t)*} = \mathbf{Y}^{(t)}\mathbf{B}_i^{(t)}$ and calculate $\hat{\mathbf{X}}_i^{(t)}$ by replacing the missing values of the original $\mathbf{X}_i^*$ matrix with the corresponding elements of $\mathbf{X}_i^{(t)*}$, whilst keeping the observed values unaltered.

5. Go to step 2, and repeat until the differences between two subsequent $\mathbf{Y}^{(t)}$ matrices becomes smaller than a certain convergence criterion.

   Note that, like before, index i indicates different observation matrices, whereas index t was used to indicate different iterations.

6 Simulation

We will assess the performance of our new approach and the GENCOM algorithm using a simulation study. In this simulation study, synthetic data is generated for several parameter settings so that the methods can be evaluated under various conditions. To assess the performance of the methods we will use the measures described in Section 4. Moreover, as the true configuration is known, we will be able to see how well the methods succeed in recovering the true structure.

6.1 Experimental design

In generating the synthetic data sets we vary a number of factors that might affect the performance of the methods. First of all, the number of $\mathbf{X}_i$ matrices, that is

the number of observations n, was chosen to be either low (10) or high (120). The case with low n corresponds to applications where we have several multidimensional scaling solutions and we want to obtain a group configuration. The high value of n corresponds to the situation in which a group of individuals is asked to evaluate objects. Secondly, although the size of the complete set of objects m was fixed at 14, the distribution of the number of objects evaluated by each individual, i.e. m_i, was varied. In particular, we considered three cases: low, medium and high variability. In the low variability case the number of objects follows a discrete uniform distribution between 4 and 6, for the medium variability the number of objects is uniform between 4 and 10, and for the high variability the number of objects is uniformly distributed between 4 and 14. Moreover, to determine which objects are observed by the ith individual m_i different numbers between 1 and 14 were drawn (without replacement). These numbers then corresponded to "observed" rows and entries of the other, i.e. the "unobserved" rows, were changed to zeros. The m_i numbers were drawn in such a way that the odds for the first four, the next six and the last four rows were: 5 to 3 to 1. Thus, the first 4 rows were 5 times as likely to be drawn than the last 4 rows. By using such a distribution we avoid the situation in which, on average, all objects are observed equally often. Instead we have a group of common objects (many observations), a group of somewhat less common objects and a group of rare objects. The absolute number of observations for each group depends of course on the number of observations n.

Thirdly, for the number of attributes used by each individual, i.e. the number of columns p_i of $\mathbf{X}_i$, we considered two cases: Low variability; number of attributes uniform between 2 and 4, and high variability; number of attributes uniform between 2 and 10. Fourthly, we considered four levels for the amount of noise added to the true (simulated) configurations. As $\mathbf{Y}_{true}$ is orthogonal its values lie between -1 and 1. We perturb the $\mathbf{Y}_{true}$ matrices by adding white noise multiplied by a factor r and we chose as values for r: $0.125, 0.25, 0.375$ and 0.5. This means that for the high noise case, i.e. $r = 0.5$, the noise follows a normal distribution with standard deviation equal to 0.5. Finally, the dimensionality of the group configuration was fixed at two, allowing graphical representations of the results.

In order to study the influence of the parameter settings we simulated for each true configuration matrix $\mathbf{Y}_{true}$, 25 $\mathbf{X}$ matrices corresponding to the same parameter settings. Thus, for one $\mathbf{Y}_{true}$ we have $25 \times 4 \times 3 \times 2 \times 2 = 1200$ simulated $\mathbf{X}$ matrices.

6.2 Data generation process

The data generation process can be summarized as follows:

1. An $m \times k$ group configuration $\mathbf{Y}_{true}$ is constructed by drawing from a standard normal distribution and then calculating an orthogonal base.

2. White noise multiplied by a factor r is added to $\mathbf{Y}_{true}$.

3. For each individual we randomly choose the number of rows, m_i, the number of columns, p_i, as well as a vector of m_i different numbers (between 1 and

m) which indicate which rows we consider to be observed by the ith individual. "Non-observed" rows are replaced by zeroes. (Later, when applying GENCOM, these zeroes will be imputed according to the GENCOM algorithm as outlined in Section 5). The resulting $m \times k$ matrix is post multiplied by a (uniform) random matrix of order $k \times p_i$. Thus, the observation matrices $\mathbf{X}_i$ are of order $m \times p_i$ and contain m_i non-zero (i.e. observed) rows.

4. The n individual $\mathbf{X}_i$ matrices are collected in one $mn \times p$ matrix $\mathbf{X}$.

5. Green and Carroll's GENCOM and our new generalized CANCOR approach are applied to the simulated observations matrices collected in $\mathbf{X}$.

6.3 Evaluation criteria and analysis

To assess the performance of our new method and the GENCOM approach we consider, as measure for the model fit, the average redundancy index $\overline{RI}_{X|Y}$ as described in Section 4.2. In addition, as the true configuration is known, we can also assess the fit of our configuration with respect to the true configuration. To do this we employ the alienation coefficient as described by Borg and Leuter (1985). The alienation coefficient, which lies between zero and one, can be interpreted as a measure of unexplained variance. We use it to compare the Euclidean distances between the rows of the true configuration, with the Euclidean distances between the rows of the retrieved configuration. The lower the alienation coefficient, the better the retrieval of the original distances, and perfect recovery is indicated by zero.

Finally, for our new method, we plot the average eigenvalues (i.e. the weighted sums of squared correlations between dimensions of the group configuration and the linear combinations) for several parameter settings, to see whether the distribution of the eigenvalues provides a good indicator for determining the dimensionality of the solution. A similar procedure cannot be used for GENCOM because, in GENCOM, each iteration involves the computation of eigenvalues and eigenvectors of a function of the imputed data matrices. As seen in step 3 of the GENCOM algorithm, the imputations depend on the choice of the dimensionality k. Consequently, the distribution of the eigenvalues depends on the chosen dimensionality and does not provide information concerning the true dimensionality.

6.4 Results

The average redundancy indices are gathered in Table 1 and the average alienation coefficients can be found in Table 2. We see that, for all conditions, our new method yields a higher fit than the GENCOM approach. The recovery of the true structure, as measured by the alienation coefficient, is also seen to be better for the new method. (Recall that the alienation coefficient is a measure of unexplained variance, the lower the value, the better the recovery of original distances).

For both methods the fit decreases if more noise is added and increases when, on average, more objects are considered (i.e. the case object variation high). Hence, when there are few missing rows in each data matrix, the fit is better. Increasing

the attribute variance does not affect the results for our method. For the GENCOM approach, however, it leads to a lower fit and poorer recovery. A higher attribute variance means that individual observation matrices consist, on average, of more columns than in the case with low attribute variance. The underlying structure remains two-dimensional. Apparently, our new method is not affected by the redundant information given in the extra columns. The GENCOM approach on the other hand, suffers considerably from the higher dimensionality of the data matrices. This is probably due to the fact that by increasing the number of attributes more elements are imputed.

In Tables 1 and 2 we see that, for the simulations with a low number of individuals (n=10), both methods yield a higher fit but poorer recovery. This can be explained by the fact that with fewer observations, it becomes easier to fit noise. The sample variances for both the redundancy index and the alienation coefficient are also larger for both methods when there are fewer observations.

The effect of the level of object variation on the recovery of the true structure is strong for both methods. More object variation yields a much better recovery of the true structure. Especially when the amount of noise is low.

A repeated measures analysis of variance was carried out to establish whether the differences for the different conditions where significant. Except for the effect of the attribute variation in the new method, all differences where significant at the 0.05 level.

In Figures 1 and 2 the average eigenvalues are plotted for different parameter settings. (As the number of attributes does not affect the solution of our method the plots corresponding to high attribute variation are not presented here.) It is immediately clear from these plots that the object variation plays a major role in the distribution of the eigenvalues. In particular, for a high object variation and a low level of noise (i.e. $r \leq 0.25$) there is a sharp drop in the value of the eigenvalues after the second eigenvalue. Thus indicating, correctly, that a two dimensional solution should be considered. If the amount of added noise increases, that is if we consider higher values for the coefficient r, the drop is less clear but, for high object variation, it remains present. When the number of observations is low, i.e. $n = 10$, the drop in the values of the eigenvalues is less clear and, for low object variation and high noise, non-existent.

7 Application to empirical data

To illustrate the generalized CANCOR method introduced in this paper we apply it in an empirical study on the positioning of political parties in The Netherlands. In March 2003 elections were held in the 12 provinces of the Netherlands for the provincial governments. A total of 28 political parties took part in the elections. Of these 28 political parties only 6 joint the elections in all 12 provinces, whereas 18 parties took part in only one province.

In order to help the voter in comparing the political programs of all parties in a particular province, several tests appeared on the Internet. One popular test was the so-called stemwijzer (vote-indicator) developed by the Instituut voor Politiek en

Table 1: Average Redundancy Indices (VB indicates our new method, GC stands for the GENCOM approach).

n=120

	Attribute Variation Low						Attribute Variation High					
Obj. Var:	Low		Med		High		Low		Med		High	
Noise	VB	GC	VB	GC	VB	GC	VB	GC	VB	GC	VB	GC
0.125	0.81	0.70	0.83	0.79	0.84	0.82	0.82	0.60	0.83	0.71	0.84	0.78
0.250	0.74	0.65	0.67	0.65	0.66	0.64	0.75	0.56	0.67	0.55	0.66	0.53
0.375	0.72	0.64	0.60	0.58	0.56	0.54	0.72	0.55	0.60	0.49	0.56	0.44
0.500	0.72	0.63	0.58	0.56	0.53	0.49	0.71	0.55	0.58	0.47	0.52	0.40

n=10

0.125	0.81	0.70	0.83	0.79	0.84	0.82	0.82	0.60	0.83	0.71	0.84	0.78
0.250	0.74	0.65	0.67	0.65	0.66	0.64	0.75	0.56	0.67	0.55	0.66	0.53
0.375	0.72	0.64	0.60	0.58	0.56	0.54	0.72	0.55	0.60	0.49	0.56	0.44
0.500	0.72	0.63	0.58	0.56	0.53	0.49	0.71	0.55	0.58	0.47	0.52	0.40

Table 2: Average Alienation Coefficients (VB indicates our new method, GC stands for the GENCOM approach).

n=120

	Attribute Variation Low						Attribute Variation High					
Obj. Var:	Low		Med		High		Low		Med		High	
Noise	VB	GC	VB	GC	VB	GC	VB	GC	VB	GC	VB	GC
0.125	0.25	0.58	0.18	0.17	0.13	0.08	0.25	0.74	0.18	0.41	0.13	0.19
0.250	0.33	0.69	0.21	0.38	0.15	0.19	0.33	0.81	0.21	0.64	0.15	0.51
0.375	0.42	0.75	0.27	0.57	0.20	0.43	0.42	0.83	0.27	0.75	0.20	0.65
0.500	0.49	0.77	0.36	0.65	0.27	0.56	0.49	0.83	0.36	0.80	0.27	0.76

n=10

0.125	0.45	0.54	0.28	0.30	0.21	0.19	0.45	0.63	0.28	0.41	0.20	0.26
0.250	0.53	0.58	0.42	0.44	0.31	0.33	0.53	0.65	0.42	0.56	0.32	0.49
0.375	0.57	0.60	0.51	0.53	0.43	0.47	0.57	0.67	0.51	0.63	0.43	0.60
0.500	0.59	0.62	0.55	0.56	0.50	0.54	0.59	0.67	0.55	0.65	0.51	0.64

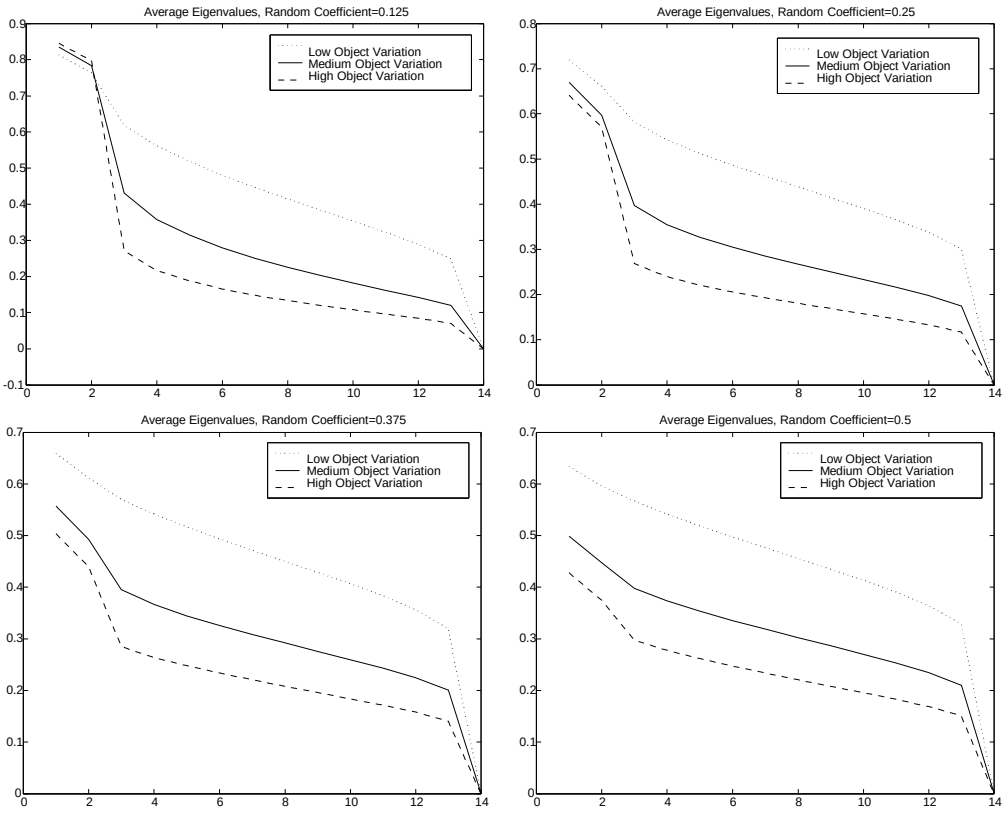

Figure 1: Average eigenvalues for n=120 and different levels of object variation and noise.

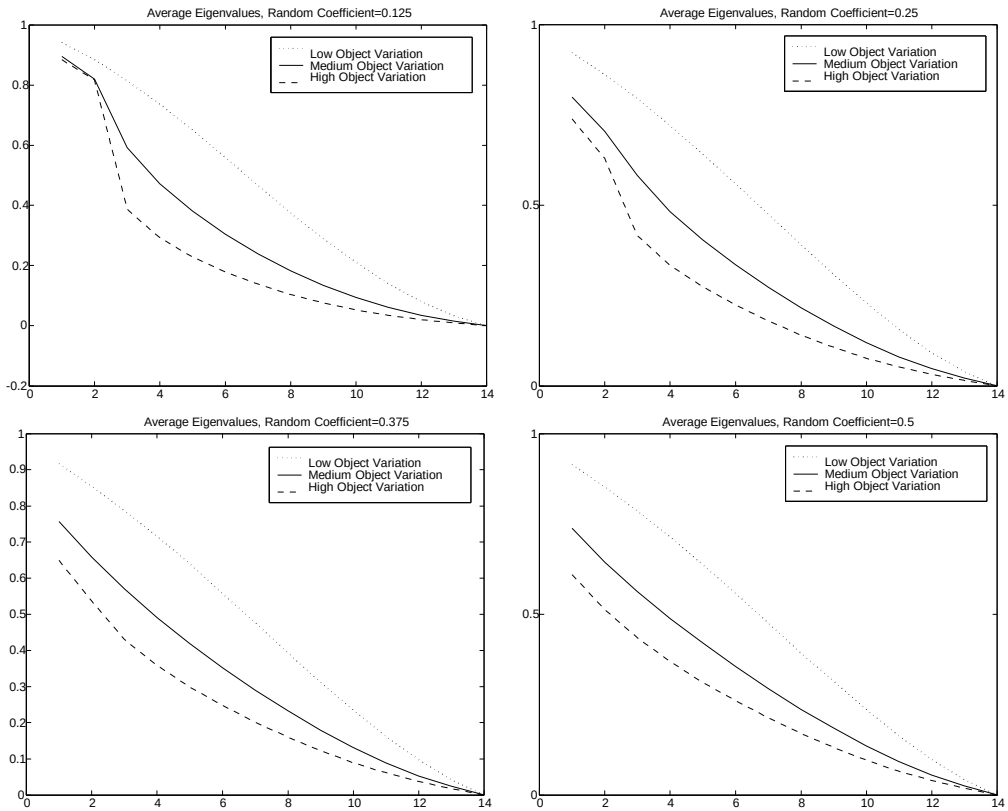

Figure 2: Average eigenvalues for n=10 and different levels of object variation and noise.

Participatie (IPP; Dutch Centre for Political Participation) in collaboration with the Documentatiecentrum Nederlandse Politieke Partijen (DNPP; Documentation centre Dutch Political Parties). The test was constructed in the following way. For each province, a set of propositions concerning relevant political issues was assembled and send to the participating political parties. Each party then reported whether they agreed, disagreed or were neutral with respect to each proposition. These propositions (ranging from 22 in one province to 31 in two others) were included in the stemwijzer website (www.stemwijzer.nl).

Individuals could access the stemwijzer website, select the appropriate province, and indicate for each of the propositions whether they agreed, disagreed, were neutral or did not have an opinion. The respondent's opinions were then compared to those of the participating parties and a summary of this comparison was provided assisting the respondent in determining which party to vote for.

We used the official party evaluations of the propositions (which can be found on the stemwijzer website) to construct, for each province, a map of the relative positions of the participating political parties. For this purpose we applied PRINCALS (principal component analysis for ordinal data, see e.g., Gifi, 1990) to the raw data. In each province, a four dimensional solution accounted for approximately 80% of the variance. Hence, we have twelve configurations each of which depicts the relative positions of a subset of the 28 political parties. The size of the subsets ranged from 8 (in Limburg) to 14 (in Noord-Holland). The average number of participating parties in each province was 10. To obtain a map of all 28 political parties, we applied the generalized CANCOR method to the twelve 4-dimensional PRINCALS solutions.

To decide on the number of dimensions of the group configuration the weighted sums of squared correlations, i.e. the λ's, do not provide a clear cut-off point. However, on the basis of the average redundancy index ($\overline{RI}_{X|Y} = 0.60$) and the interpretability of the group configuration, we selected the two dimensional solution. Adding a third dimension led to a relatively small (0.13) increase in the average redundancy.

It is easily verified that rotation of $\mathbf{Y}$ and $\mathbf{A}_i$ (for $i = 1, ..., n$) over the same angle, does not affect the value of objective function (6). Hence, in order to simplify its interpretation, we rotated the two-dimensional group configuration to simple structure using Kaiser's varimax criterion. The rotated configuration can be found in Figure 3.

To interpret the group configuration we will first focus on parties that joined the elections in at least 6 provinces. These parties, which are circled in Figure 3, also participated in the national elections and their relative positions in the Dutch political landscape are relatively well understood. We see that these 9 parties lie on a semi-circle. Starting from the upper-left we can move counterclockwise from the left-wing parties, Groen Links and SP, to the right-wing party LPF. Hence, the first dimension describes the traditional distinction between left-wing and right-wing parties. The second dimension separates parties that can be considered to be more traditional and conservative, in particular with respect to issues concerning political and governmental innovations, from the, in that respect, more progressive political parties. All religious parties (Christen Unie, SGP, and CDA) are located at the low end (traditional) of dimension 2. The relative positions of these nine parties are

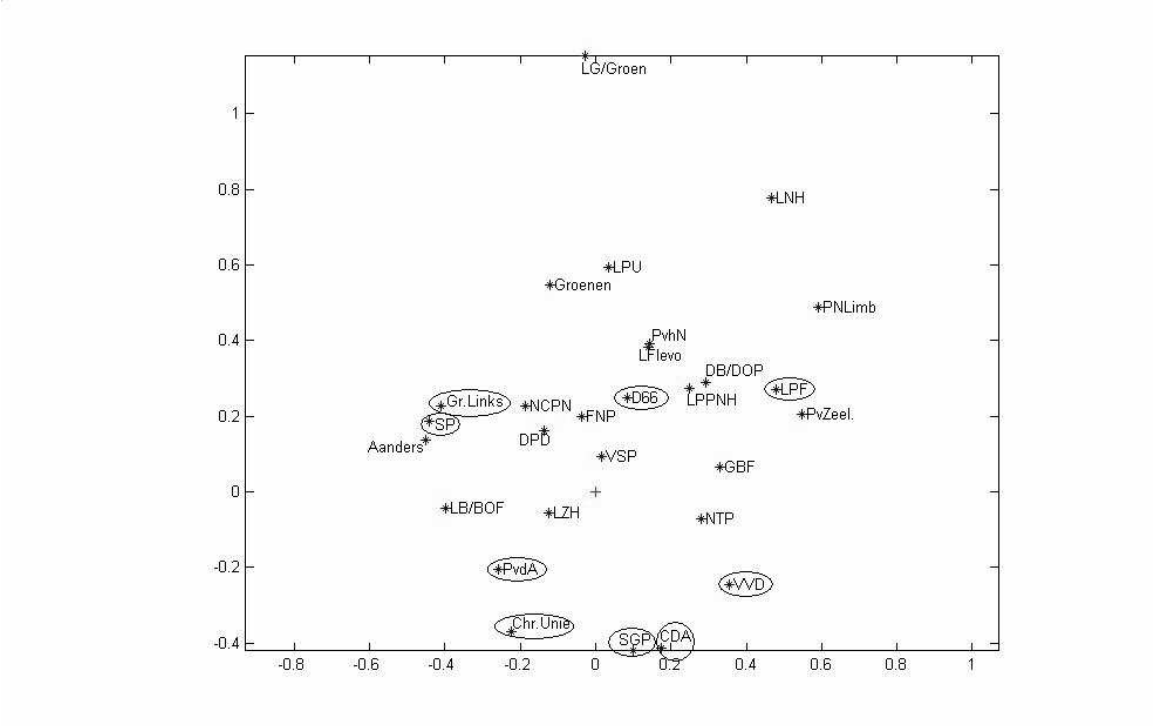

Figure 3: Group configuration (after varimax rotation) obtained using the generalized CANCOR approach.

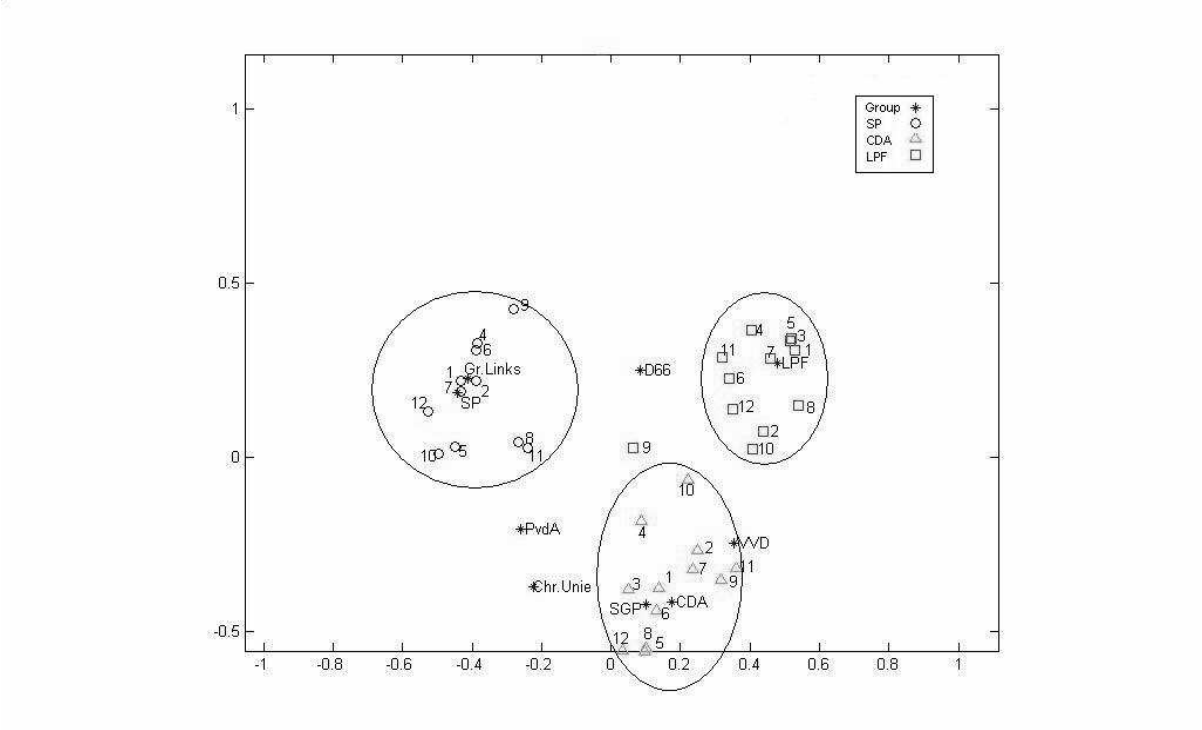

Figure 4: Points corresponding to the parties SP, CDA and LPF in the individual configurations.

similar to those obtained by Groenen (2003) in his mapping of political parties based on their political programs for the Dutch national elections of January 2003.

The group configuration makes it possible to see the relative positions of local political parties with respect to the national ones. This is done in nearly all cases on the bases of only one observation for the local parties. Furthermore, the group configuration of our CANCOR procedure reveals the relative positions of many political parties that did not overlap in the provinces they joint in the actual elections. In Figure 3 we see that the local parties are located relatively close to the progressive parties and are far from the traditional parties. Hence, the second dimension also separates the regional parties from (some of the) national parties.

The squared correlations between dimensions of the individual configuration and the group configuration (i.e. the correlations between the columns of $\mathbf{Y}_i$ and $\mathbf{Y}$) are given in Table 3. These squared correlations give an indication of how well the individual configurations are represented by the group configuration. We see that, in general, these squared correlations are quite high. This indicates that the positions of the political parties are quite stable over the provinces. This is illustrated in Figure 4, where we have plotted the points that correspond to the LPF, CDA and the SP for each individual configuration, together with the points corresponding to the nine large parties in the group configuration.

It is difficult to discern, without further research, why the squared correlations for Zuid-Holland (dimension 2) and Noord-Brabant are relatively low. However, the

Table 3: Squared correlations

		CANCOR		GENCOM	
		dim 1	dim 2	dim 1	dim 2
1)	Friesland	0.97	0.98	0.68	0.88
2)	Groningen	0.94	0.89	0.97	0.82
3)	Drente	0.92	0.87	0.60	0.83
4)	Overijssel	0.89	0.77	0.59	0.73
5)	Gelderland	0.96	0.92	0.88	0.84
6)	Flevoland	0.97	0.89	0.73	0.81
7)	Utrecht	1.00	0.94	0.77	0.83
8)	Noord-Holland	0.88	0.88	0.95	0.91
9)	Zuid-Holland	0.94	0.17	0.61	0.29
10)	Zeeland	0.88	0.77	0.52	0.87
11)	Noord-Brabant	0.45	0.69	0.51	0.51
12)	Limburg	0.96	0.88	0.79	0.94

discrepancy suggests that in these provinces, the relative positions of the national parties with respect to each other are quite different from the structure in other provinces. The poor correspondence between the individual configuration of Noord-Brabant with the group configuration (as indicated by the low squared correlations) is illustrated in Figure 5, where we have plotted the individual configuration together with the appropriate points in the group configuration. In Figure 6 a similar plot is presented for the, well represented, province Utrecht.

GENCOM We also applied GENCOM to the political party data. We chose the algorithm to terminate when the sum of squared differences between two subsequent group configurations became smaller than 2.2×10^{-16} (computer accuracy). This occurred after 740 iterations. In order to simplify the comparison between the two group configurations we rotated the GENCOM solution in such a way that the sum of squared differences between the GENCOM and the CANCOR solution was at a minimum. In other words, we applied Procrustes rotation to the GENCOM solution. The resulting two-dimensional group configuration is presented in Figure 7.

The differences between the configurations obtained using GENCOM and CANCOR are relatively small. In particular, the positions of the large political parties (i.e. the ones that joined in more than 9 provinces) are quite similar in both plots. The positions of some of the regional parties, however, differ substantially between the GENCOM and CANCOR solutions. The average redundancy index for the two-dimensional GENCOM solution is 0.53. The squared correlations between dimensions of the individual configurations with the group configuration, both for the CANCOR and the GENCOM solutions, can be found in Table 3. Hence, the fit, both in terms of the average redundancy index– 0.60 for CANCOR versus 0.53 for GENCOM– and the average squared correlations, is better in our non-iterative CANCOR approach.

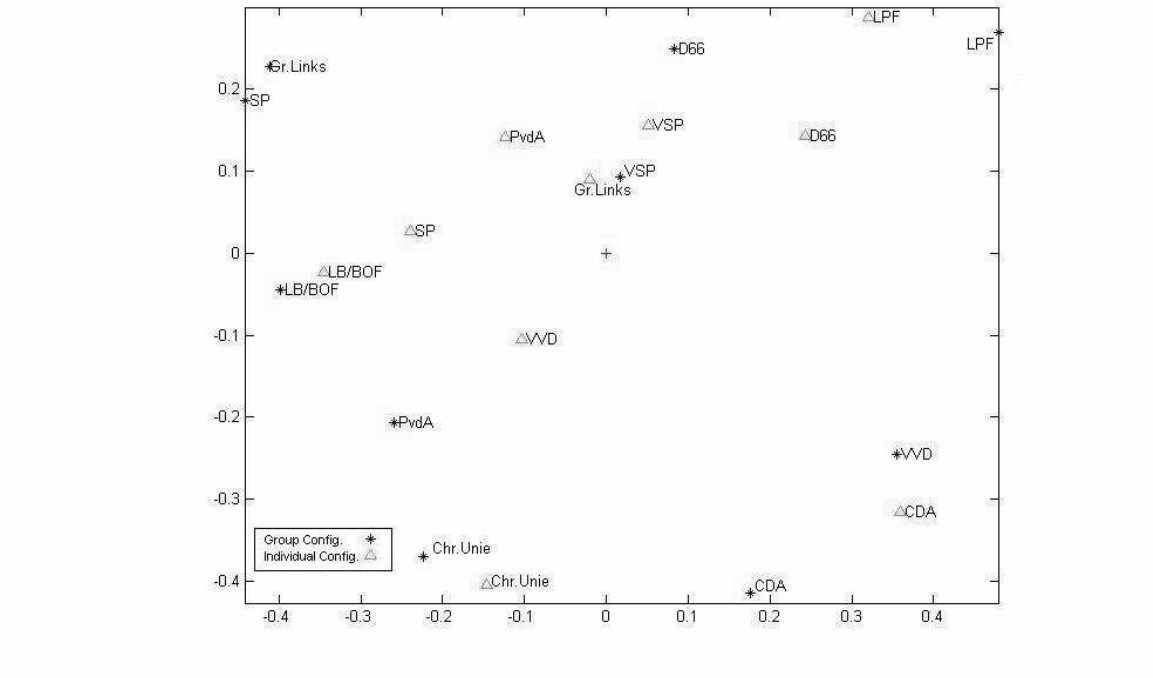

Figure 5: Joint plot of the individual configuration of Noord-Brabant and the group configuration.

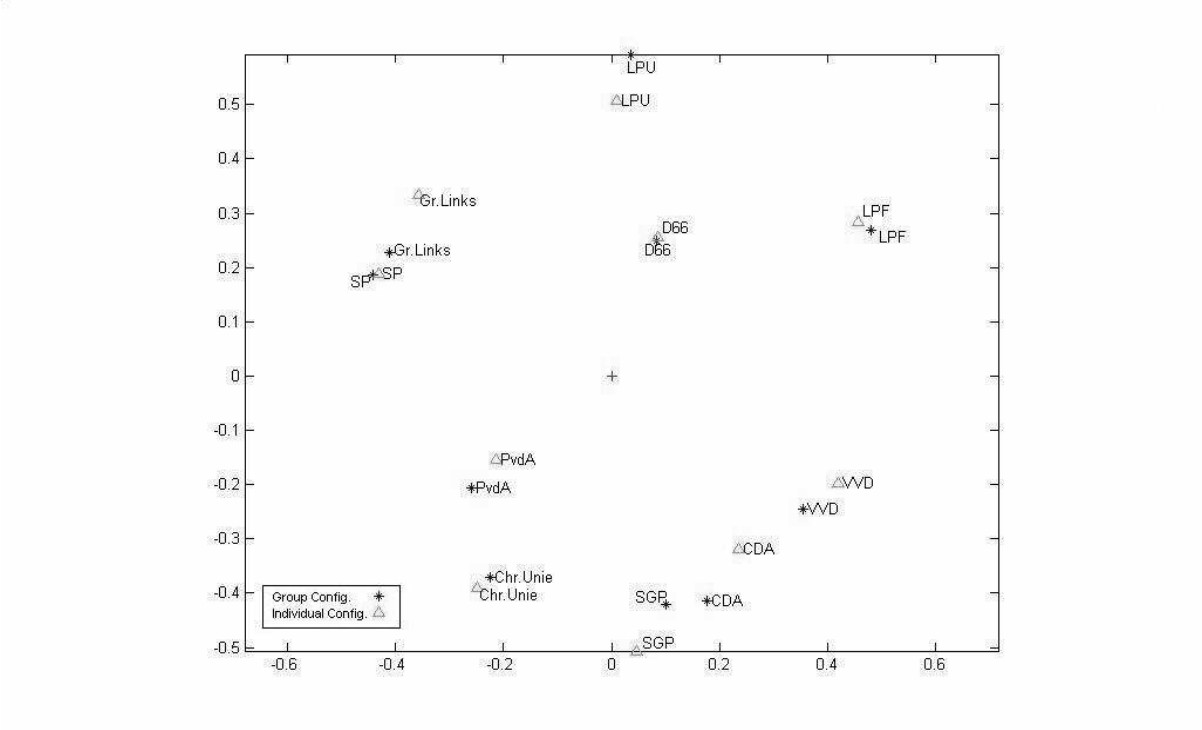

Figure 6: Joint plot of the individual configuration of Utrecht and the group configuration.

8 Discussion

In this paper we offered a method that makes it possible to apply generalized canonical correlation analysis to matrices of different row and column orders. This new method derives a group configuration for a set of objects based on observation matrices of (potentially) different row and column order. In contrast to GENCOM, an existing generalization proposed by Green and Carroll (1988) that makes it possible to apply CANCOR in the presence of missing data, our new method does not require any type of iteration. The solution is obtained immediately by means of an eigenvalue decomposition.

A simulation study was used to assess the performance of our new method and to compare it to the GENCOM approach. We found that, both with respect to the fit as well as the recovery of the true structure, our method performed better than the GENCOM approach. This was further illustrated in the empirical example concerning the political landscape in the Netherlands.

There are many fields in which our new method can be applied. It can, for example, be used to obtain a configuration for a set of objects on the basis of several configurations of subsets. Our analysis of the political landscape in the Netherlands based on provincial data is an example of such an application. Other possible applications concern the analysis of data in which individuals evaluate a set of objects (e.g. products, brands, political parties) with which they are familiar, using a set of

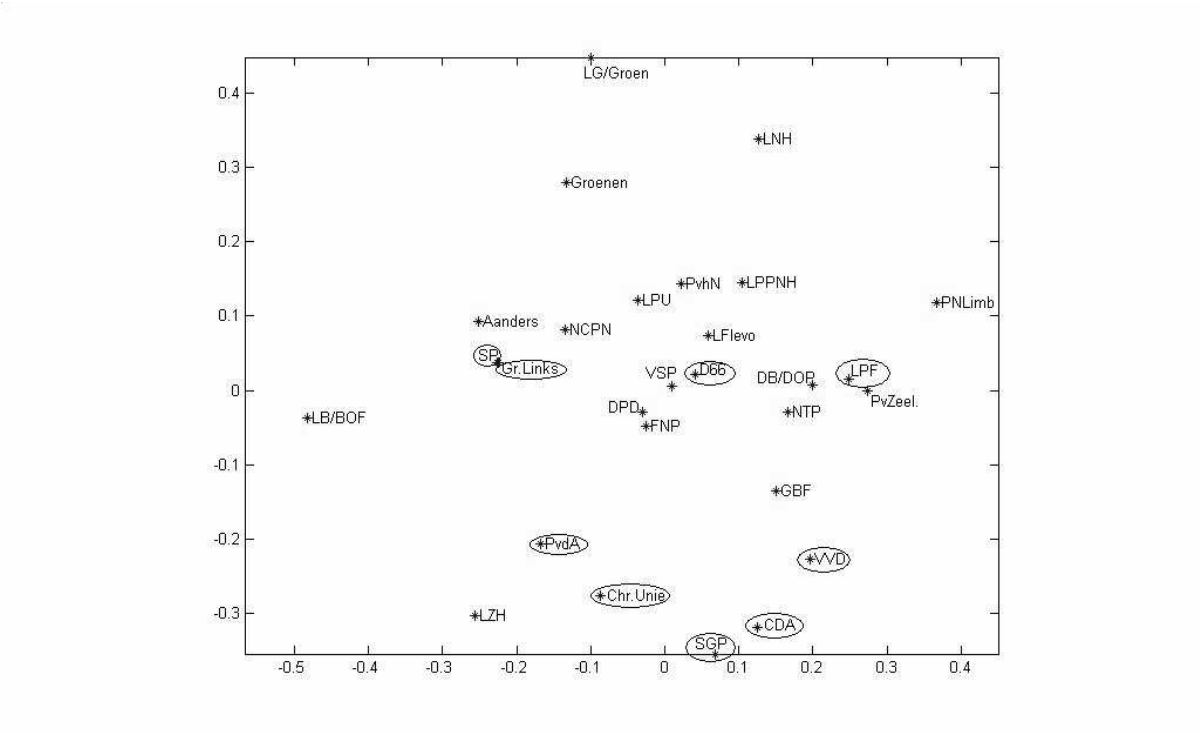

Figure 7: Group configuration (after rotation) obtained using the GENCOM algorithm.

attributes that they deem relevant. Hence, both the object and attribute sets are individual specific. Problems concerning a-priori fixed sets of objects and attributes, such as the familiarity and relevance of the objects and attributes to all individuals, are thus avoided. In addition, the task for the individuals will be easier when they only need to judge objects that they are familiar with, using relevant criteria.

An important problem that one encounters when applying the generalized CANCOR approach introduced in this paper concerns the dimensionality of the group configuration. That is, the choice of the number of dimensions that are needed to adequately summarize the data. In our simulation study we showed that the sum of weighted squared correlations, i.e. the eigenvalues, provide a good indication concerning the true dimensionality when the level of noise is low and the subsets of objects used by the individuals are relatively large. When there is no clear drop in the values of the eigenvalues, a choice can be made on the basis of the average redundancy index and the interpretability of the solution. Further research is needed to obtain a more objective measure to determine the number of dimensions.

Throughout the paper we have restricted ourselves to the generalized canonical correlation analysis approach commonly referred to as CANCOR. There are, however, other multivariate methods that can be used to obtain a group configuration based on several data sets. Steenkamp, Van Trijp, and Ten Berge (1994), in their analysis of matrices of equal row and different column order, describe empirical results of three different methods: Generalized Procrustes analysis (Gower, 1975, Ten Berge,

1977), INDSCAL (Carroll and Chang, 1970), and CANCOR. They find that the three methods perform about equally well. In addition, Commandeur (1991) and Ten Berge, Kiers and Commandeur (1993) offer generalizations of Procrustes analysis that allow the analysis of several data sets with missing rows and columns. It would be interesting to see whether the similar results obtained when applying generalized Procrustes analysis and CANCOR to matrices of different column order carry over to the more general case where the data matrices have different row and column orders.

The results obtained in both the simulation study and the empirical illustration, suggest that the method introduced in this paper can be of great practical value in a wide range of applications. Its flexibility can be used to allow individuals a much greater freedom in evaluating sets of objects and obtaining a group configuration based on these evaluations.

A Derivation of the generalized CANCOR solution

Recall the objective function (6):

$$\min_{\mathbf{A},\mathbf{Y}} \ \phi = \operatorname{trace} \sum_{i=1}^{n} \left(\mathbf{K}_i\mathbf{Y} - \mathbf{K}_i\mathbf{X}_i\mathbf{A}_i\right)' \left(\mathbf{K}_i\mathbf{Y} - \mathbf{K}_i\mathbf{X}_i\mathbf{A}_i\right). \tag{15}$$

$$s.t.\ \mathbf{Y}'\mathbf{K}\mathbf{Y} = n\mathbf{I}_k.$$

where $\mathbf{K} = \sum_{i=1}^{n} \mathbf{K}_i$ and we have discarded the superscribed $*$'s for ease of notation. Expanding the bracketed expressions yields

$$\begin{aligned} \phi &= \operatorname{trace} \mathbf{Y}'\mathbf{K}\mathbf{Y} + \sum_{i=1}^{n} \operatorname{trace} \mathbf{A}_i'\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i\mathbf{A}_i - 2\sum_{i=1}^{n} \operatorname{trace} \mathbf{Y}'\mathbf{K}_i\mathbf{X}_i\mathbf{A}_i \\ &= \operatorname{trace} \mathbf{Y}'\mathbf{K}\mathbf{Y} + \operatorname{trace} \mathbf{A}'\mathbf{D}\mathbf{A} - 2\operatorname{trace} \mathbf{Y}'\mathbf{X}\mathbf{A}, \end{aligned}$$

where $\mathbf{X} \equiv \begin{pmatrix} \mathbf{K}_1\mathbf{X}_1 & \mathbf{K}_2\mathbf{X}_2 & \cdots & \mathbf{K}_n\mathbf{X}_n \end{pmatrix}$, $\mathbf{A} \equiv \begin{pmatrix} \mathbf{A}_1' & \mathbf{A}_2' & \cdots & \mathbf{A}_n' \end{pmatrix}'$ and

$$\mathbf{D} \equiv \begin{pmatrix} \mathbf{X}_1'\mathbf{K}_1\mathbf{X}_1 & & & \\ & \mathbf{X}_2'\mathbf{K}_2\mathbf{X}_2 & & \\ & & \ddots & \\ & & & \mathbf{X}_n'\mathbf{K}_n\mathbf{X}_n \end{pmatrix}.$$

Minimization of ϕ is equivalent to maximization of $-\phi$ and hence (15) is equivalent to

$$\max_{\mathbf{A},\mathbf{Y}} \ \phi^* = 2\operatorname{trace} \mathbf{Y}'\mathbf{X}\mathbf{A} - \operatorname{trace} \mathbf{A}'\mathbf{D}\mathbf{A} - \operatorname{trace} \mathbf{Y}'\mathbf{K}\mathbf{Y}$$

$$s.t.\ \mathbf{Y}'\mathbf{K}\mathbf{Y} = n\mathbf{I}_k.$$

Inserting the constraint gives as Lagrangian function

$$\psi = 2\operatorname{trace} \mathbf{Y}'\mathbf{X}\mathbf{A} - \operatorname{trace} \mathbf{A}'\mathbf{D}\mathbf{A} - nk - \operatorname{trace} \mathbf{L}(\mathbf{Y}'\mathbf{K}\mathbf{Y} - n\mathbf{I}_k)$$

where $\mathbf{L}$ is the $k \times k$ matrix of Lagrange multipliers. Note that, as the restrictions are symmetric, the matrix of Lagrange multipliers is symmetric, i.e. $\mathbf{L} = \mathbf{L}'$. Taking derivatives with respect to $\mathbf{A}$ and $\mathbf{Y}$ yields as first order conditions

$$d\mathbf{A} : 2 \operatorname{trace} \mathbf{Y}'\mathbf{X} d\mathbf{A} - 2 \operatorname{trace} \mathbf{A}'\mathbf{D} d\mathbf{A} = 0 \tag{16}$$

$$d\mathbf{Y} : 2 \operatorname{trace} \mathbf{A}'\mathbf{X}' d\mathbf{Y} - 2 \operatorname{trace} \mathbf{L}\mathbf{Y}'\mathbf{K} d\mathbf{Y} = 0 \tag{17}$$

(for a treatment of matrix differentials see, e.g., Magnus and Neudecker, 1999). From (16) we get

$$\mathbf{X}'\mathbf{Y} = \mathbf{D}\mathbf{A}.$$

Hence

$$\mathbf{A}'\mathbf{X}'\mathbf{Y} = \mathbf{A}'\mathbf{D}\mathbf{A}, \tag{18}$$

and, assuming that the matrices $\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i$ are nonsingular for all i,

$$\mathbf{A} = \mathbf{D}^{-1}\mathbf{X}'\mathbf{Y}. \tag{19}$$

Using the definitions for $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{X}$, we see that

$$\mathbf{A}_i = \left(\mathbf{X}_i'\mathbf{K}_i\mathbf{X}_i\right)^{-1} \mathbf{X}_i\mathbf{K}_i\mathbf{Y}.$$

Due to the symmetry of $\mathbf{K}$ and $\mathbf{L}$ the first-order condition in (17) can be expressed as

$$\mathbf{X}\mathbf{A} = \mathbf{K}\mathbf{Y}\mathbf{L}. \tag{20}$$

Inserting the expression for $\mathbf{A}$ yields

$$\mathbf{X}\mathbf{D}^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{K}\mathbf{Y}\mathbf{L}. \tag{21}$$

Premultiplying this by $\frac{1}{\sqrt{n}}\mathbf{K}^{-\frac{1}{2}}$ gives

$$\frac{1}{\sqrt{n}}\mathbf{K}^{-\frac{1}{2}}\mathbf{X}\mathbf{D}^{-1}\mathbf{X}'\mathbf{K}^{-\frac{1}{2}}\mathbf{K}^{\frac{1}{2}}\mathbf{Y} = \frac{1}{\sqrt{n}}\mathbf{K}^{\frac{1}{2}}\mathbf{Y}\mathbf{L}.$$

Let

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$$

where $\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_k$ and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues. Then

$$\frac{1}{\sqrt{n}}\mathbf{K}^{-\frac{1}{2}}\mathbf{X}\mathbf{D}^{-1}\mathbf{X}'\mathbf{K}^{-\frac{1}{2}}\mathbf{K}^{\frac{1}{2}}\mathbf{Y}\mathbf{U} = \frac{1}{\sqrt{n}}\mathbf{K}^{\frac{1}{2}}\mathbf{Y}\mathbf{U}\mathbf{\Lambda}$$

and, without loss of generality, we can express (21) as

$$\mathbf{K}^{-\frac{1}{2}}\mathbf{X}\mathbf{D}^{-1}\mathbf{X}'\mathbf{K}^{-\frac{1}{2}}\mathbf{Y}^* = \mathbf{Y}^*\mathbf{\Lambda},$$

where $\mathbf{Y}^* = \frac{1}{\sqrt{n}}\mathbf{K}^{\frac{1}{2}}\mathbf{Y}\mathbf{U}$ so that $\mathbf{Y}^{*'}\mathbf{Y}^* = \mathbf{I}$. Hence, $\mathbf{Y}^*$ is a matrix of eigenvectors corresponding to a matrix of eigenvalues $\mathbf{\Lambda}$.

From (20) it follows that $\mathbf{L} = \frac{1}{n}\mathbf{Y}'\mathbf{X}\mathbf{A}$. Inserting this and using (18) in the objective function ϕ^* yields

$$\max_{\mathbf{A},\mathbf{Y}} \phi^* = n \operatorname{trace} \Lambda - nk.$$

Thus, in order to obtain a maximum for ϕ^* we must select the k largest eigenvalues of $\mathbf{K}^{-\frac{1}{2}}\mathbf{X}\mathbf{D}^{-1}\mathbf{X}'\mathbf{K}^{-\frac{1}{2}}$ and corresponding eigenvectors.

References

Borg, I. and Leutner, D. (1985). Measuring the similarity between MDS configurations. *Multivariate Behavioral Research*, 20:325–334.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the American Psychological Association*, pages 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*, 35:283–319.

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

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley, Chichester.

Gower, J. (1975). Generalized procrustes analysis. *Psychometrika*, 40:33–51.

Green, P. E. and Carroll, J. D. (1988). A simple procedure for finding a composite of several multidimensional scaling solutions. *Journal of the Academy of Marketing Science*, 16:25–35.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press, London.

Groenen, P. (2003). *Dynamische Meerdimensionale Schaling: Statistiek Op de Kaart*. ERIM Inaugural Addresses Research in Management Series, Rotterdam.

Magnus, J. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, Chichester.

Steenkamp, J.-B. E. M., Van Trijp, H. C. M., and Ten Berge, J. M. F. (1994). Perceptual mapping based on idiosyncratic sets of attributes. *Journal of Marketing Research*, 31:15–27.

Stewart, D. and Love, W. (1968). A general canonical correlation index. *Psychological Bulletin*, 70:160–163.

Ten Berge, J. M. F. (1977). Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42:267–276.

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