RR 94-01

FITTING A DISTANCE MODEL TO HOMOGENEOUS SUBSETS OF VARIABLES: POINTS OF VIEW ANALYSIS OF CATEGORICAL DATA

Jacqueline J. Meulman

Department of Data Theory

University of Leiden

FITTING A DISTANCE MODEL TO HOMOGENEOUS SUBSETS OF VARIABLES: POINTS OF VIEW ANALYSIS OF CATEGORICAL DATA

Abstract: An approach is presented for analyzing a heterogeneous set of categorical variables assumed to form a limited number of homogeneous subsets. The variables generate a particular set of proximities between the objects in the data matrix, and the objective of the analysis is to represent the objects in low-dimensional Euclidean spaces, where the distances approximate these proximities. A least squares loss function is minimized that involves three major components: a) the partitioning of the heterogeneous variables into homogeneous subsets; b) the optimal quantification of the categories of the variables, and c) the representation of the objects through multiple multidimensional scaling tasks performed simultaneously. An important aspect from an algorithmic point of view is in the use of majorization. The behavior of the procedure is studied through an analysis of a more-dimensional data structure with known properties. Permutation tests are used to evaluate whether there is evidence of communality between different points of view, and between variables and subset(s) to which they have not been assigned. Special attention is given to the local minima problem, resulting from unfortunate initial assignments of variables to subsets. The use of the procedure is demonstrated by a typical example of possible application, i.e., the analysis of categorical data obtained in a free-sort task, and the stability is studied through a Jackknife analysis. The results of a points of view analysis are contrasted with those from standard homogeneity analysis.

Keywords: Categorical data; Multidimensional scaling; Homogeneity analysis; Optimal scaling; Majorization; Clustering of variables.

1. Introduction

Multiple sources giving proximity information on objects, to be used in a multidimensional scaling (MDS) analysis, may be homogeneous within subsets and heterogeneous between subsets of sources. Meulman and Verboon (1993) proposed an integrated approach for the analysis of such proximity data, called points of view analysis. This approach can be viewed as a streamlined version of the original Tucker and Messick (1963) procedure that consisted of computing a number of linear combinations of the proximity data (by a principal components analysis), and the subsequent fitting of distances in low-dimensional Euclidean spaces to the component scores, after a rotation giving some simple structure for the weights.

Meulman and Verboon (1993) integrated these separate steps by minimizing a single least squares loss function, defined on distances. The aim is to obtain different multidimensional object spaces for different subsets of sources that are empirically derived, with each displaying a particular viewpoint about the object interrelationships. The distances between the objects in the multidimensional spaces should resemble the proximity data for each subset of sources as closely as possible, and provide a parsimonious display of the objects in a limited number of low-dimensional points of view, say $r \geq 2$.

In the present paper, a special variety of points of view analysis is developed. The data to be analyzed are observations of N objects on M categorical variables, collected in an $N \times M$ data matrix $\mathbf{Z}$. The m-th column in $\mathbf{Z}$ contains the categorical variable $\mathbf{z}_m$, and will be associated with the indicator matrix $\mathbf{G}_m$, with N rows and k_m columns, where k_m denotes the number of different categories for variable $\mathbf{z}_m$. The elements of $\mathbf{G}_m$ are defined as follows: $g_{(m)il} = 1$ if $z_{im} = l$, and $g_{(m)il} = 0$ otherwise, $l = 1,\ldots,k_m$, $i = 1,\ldots,N$. A popular technique for analyzing this type of data is homogeneity analysis (Gifi 1990), also known as multiple correspondence analysis (Greenacre 1984), or dual scaling (Nishisato 1980, 1994). If the variables are heterogeneous, many dimensions may be required in the object space. The points of view analysis proposed here can be considered an alternative approach to accommodate heterogeneous categorical variables, by clustering them into homogeneous subsets. Each categorical variable contains proximity information on the objects; this information is processed through a series of multidimensional scaling (MDS) tasks, defined on Euclidean distances between the objects .

2. Method

Prior to describing the details of the method, some additional notation is necessary. The MDS task is to represent the objects in Euclidean spaces $\mathbf{X}_s$, $s = 1,...,r$, with assumed dimensionality p_s. The squared distance between a pair of objects $\{i,j\}$ in the s-th space is defined by

$$d_{ij}^2(\mathbf{X}_s) = (\mathbf{e}_i - \mathbf{e}_j)'\mathbf{X}_s\mathbf{X}_s'(\mathbf{e}_i - \mathbf{e}_j),$$

where $\mathbf{e}_i$ is the i-th column of the $N \times N$ identity matrix $\mathbf{I}$. It is convenient to define the squared distance function $D^2(\cdot)$ that maps coordinates, $\mathbf{X}_s$, into squared distances as:

$$D^2(\mathbf{X}_s) = \mathbf{a1}' + \mathbf{1a}' - 2\mathbf{X}_s\mathbf{X}_s',$$

where $\mathbf{a} = \text{vecdiag}(\mathbf{X}_s\mathbf{X}_s')$ is the N–vector containing the diagonal elements of $\mathbf{X}_s\mathbf{X}_s'$, and $\mathbf{1}$ is the N–vector of all 1's. The distances in $\mathbf{X}_s$ approximate the available proximities as closely as possible; the proximity information is derived as follows. The indicator matrix $\mathbf{G}_m$ is associated with a $k_m \times q_m$ matrix $\mathbf{Y}_m$, with $q_m \leq k_m$, containing q_m-dimensional quantifications for the categories in $\mathbf{z}_m$, where the dimensionality q_m may be chosen independently from the dimensionality of $\mathbf{X}_s$. Using $\mathbf{G}_m$ and $\mathbf{Y}_m$, an $N \times N$ matrix is formed, containing distances according to the m-th variable; the corresponding matrix with squared distances is given by

$$D^2(\mathbf{G}_m\mathbf{Y}_m) = \mathbf{c1}' + \mathbf{1c}' - 2\mathbf{G}_m\mathbf{Y}_m\mathbf{Y}_m'\mathbf{G}_m',$$

where $\mathbf{c} = \text{vecdiag}(\mathbf{G}_m\mathbf{Y}_m\mathbf{Y}_m'\mathbf{G}_m')$.

A salient feature of the method proposed, compared to an ad hoc approach that would perform a number of independent multidimensional scaling tasks, is in the assignment of variables to different subsets. To arrive at r subsets of variables, the index set $\{1,...,M\}$ is partitioned into subsets J_s, $s = 1,...,r$. Each variable is associated with a row of a matrix $\mathbf{A} = \{a_{ms}\}$, of order $M \times r$, containing weights that express the influence of each variable in the overall analysis, and reflect the particular structure of each variable belonging to one and only one subset. Specifically, $\mathbf{A}$ can be written in the form $\mathbf{A} = \mathbf{WB}$, where $\mathbf{W}$ is a diagonal matrix of order $M \times M$, containing a weight w_{mm} in the main diagonal that will be associated with each $D(\mathbf{G}_m\mathbf{Y}_m)$, and $\mathbf{B}$ is the group structure matrix; it is a 1/0 indicator matrix, of order $M \times r$, that assigns each $\mathbf{G}_m$ to one of the r subsets. Having defined the basic elements of analysis, the least squares loss function to be minimized can be written as

$$\text{STRESS}(\mathbf{A}; \mathbf{Y}_1,\ldots,\mathbf{Y}_M; \mathbf{X}_1,\ldots,\mathbf{X}_r) = (2NM)^{-1} \sum_{s=1}^{r} \sum_{m \in J_s} \|a_{ms}D(\mathbf{G}_m\mathbf{Y}_m) - D(\mathbf{X}_s)\|^2, \quad (1)$$

where $\|\cdot\|^2$ indicates the sum of squared differences. Particular constraints are necessary in the minimization of (1) over $\mathbf{A}$, $\{\mathbf{Y}_1,\ldots,\mathbf{Y}_M\}$, and $\{\mathbf{X}_1,\ldots,\mathbf{X}_r\}$. Minimization is over $\mathbf{A} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ denotes the set of all restricted weight matrices, satisfying $\mathbf{A} = \mathbf{WB}$. To avoid a perfect but trivial solution (by setting $\mathbf{Y}_m = \mathbf{0}$ and $\mathbf{X}_s = \mathbf{0}$), we require without loss of generality that $\|\mathbf{X}_s\|^2 = 1$. In addition, $\mathbf{X}_s$ and $\mathbf{G}_m\mathbf{Y}_m$ will be column-centered, $\mathbf{1}'\mathbf{X}_s = \mathbf{1}'\mathbf{G}_m\mathbf{Y}_m = \mathbf{0}$, and the quantifications $\mathbf{Y}_m$ are conveniently normalized such that $\|\mathbf{G}_m\mathbf{Y}_m\|^2 = 1$. By this choice of normalization the stress value can be re-expressed by $1 - M^{-1}\|\mathbf{A}\|^2$, so $M^{-1}\|\mathbf{A}\|^2$ is a goodness-of-fit measure. Due to the choice of normalization, the elements of $\mathbf{A}$ are congruence coefficients (Tucker 1951) between the proximities derived from $\mathbf{G}_m\mathbf{Y}_m$ and the distances in $\mathbf{X}_s$.

3. Background

If only one point of view is considered ($r = 1$), the method discussed in this paper encompasses features of both classical homogeneity analysis, as in the HOMALS procedure (Gifi 1990), and the distance approach to homogeneity analysis (Meulman 1986). These methods will be reviewed very briefly, to give some background to the new method. The loss function for classical homogeneity analysis is written as

$$\text{STRIFE}(\mathbf{Y}_1,\ldots,\mathbf{Y}_M; \mathbf{X}) = M^{-1} \sum_{m=1}^{M} \|\mathbf{G}_m\mathbf{Y}_m - \mathbf{X}\|^2,$$

and is minimized over $\mathbf{X}$ and $\mathbf{Y}_m$, all of dimensionality p. Each quantified variable $\mathbf{G}_m\mathbf{Y}_m$ should resemble as closely as possible a single $\mathbf{X}$ that functions as a comparison set. Normalization constraints are required to avoid the trivial result $\mathbf{X} = \mathbf{0}$ and $\mathbf{Y}_m = \mathbf{0}$, $m = 1,\ldots,M$. In the standard HOMALS analysis, constraints are introduced on $\mathbf{X}$, $\mathbf{X}'\mathbf{X} = \mathbf{I}$ and $\mathbf{1}'\mathbf{X} = \mathbf{0}$, and the optimal quantifications for given $\mathbf{X}$ are found as $\hat{\mathbf{Y}}_m = (\mathbf{G}_m'\mathbf{G}_m)^{-1}\mathbf{G}_m'\mathbf{X}$.

In terms of distance representation, it is shown in Meulman (1986) that the optimal **X** is a normalized version of the configuration of object points that approximate the chi-squared distances in each space $\mathbf{G}_m$ simultaneously through the use of the classical multidimensional scaling procedure defined on inner products of the object coordinates, originating from Torgerson (1958) and Gower (1966). In contrast, a least squares distance approach would not define a loss function on inner products, but instead on the distances directly, and minimize

$$\text{STRESS}(\mathbf{X}) = (2NM)^{-1} \sum_{m=1}^{M} \|\mathrm{D}(\mathbf{G}_m(\mathbf{G}_m'\mathbf{G}_m)^{-1/2}) - \mathrm{D}(\mathbf{X})\|^2, \tag{2}$$

where $\mathrm{D}(\mathbf{G}_m(\mathbf{G}_m'\mathbf{G}_m)^{-1/2})$ is the matrix containing the chi-squared distances between the rows of $\mathbf{G}_m$. In the distance approach, the configuration **X** is usually positioned in principal axes orientation: the dimensions are rotated towards orthogonality and display differential saliences (measured by the sums of squares of the object coordinates). It should be noted that the minimization of (2) is equivalent to the minimization of

$$\text{STRESS}(\mathbf{Y}_1,\ldots,\mathbf{Y}_M;\, \mathbf{X}) = (2NM)^{-1} \sum_{m=1}^{M} \|\mathrm{D}(\mathbf{G}_m\mathbf{Y}_m) - \mathrm{D}(\mathbf{X})\|^2,$$

provided that the dimensionality, q_m, of each $\mathbf{Y}_m$ is set equal to k_m, and weighted orthonormality constraints are imposed on the quantifications: $\mathbf{Y}_m'\mathbf{G}_m'\mathbf{G}_m\mathbf{Y}_m = \mathbf{I}$; since distances are invariant under rotation, $\mathrm{D}(\mathbf{G}_m\mathbf{Y}_m) = \mathrm{D}(\mathbf{G}_m(\mathbf{G}_m'\mathbf{G}_m)^{-1/2})$.

The method proposed here resembles classical homogeneity analysis by giving sets of low-dimensional quantifications $\mathbf{G}_m\mathbf{Y}_m$ and follows the distance approach by direct approximation of distances, derived from each variable separately (after applying the appropriate weight from **A**). The most important reason to choose for an objective function defined on distances is that a least squares distance approximation compared to the use of inner products usually needs fewer dimensions to represent the proximity information satisfactorily. This argument holds even stronger when the data contain a dominant first dimension, which in a standard homogeneity analysis would result in the Guttman effect, giving subsequent dimensions that are polynomial functions of the first one.

It should be noted, however, that a remarkable property of a standard HOMALS solution no longer applies in the distance approach. In a standard HOMALS solution, with $\mathbf{X}'\mathbf{X} = \mathbf{I}$, the optimal p-dimensional category quantifications in $\mathbf{Y}_m$ give coordinates for category points that are in the centroid of the objects in the same category according to variable m. When distances are approximated, this so-called centroid principle no longer holds. Considering the method as a technique that quantifies qualitative variables and using the property that distances are invariant under rotation, the optimal $\hat{\mathbf{Y}}_m$ may be rotated towards principal axes orientation to show differential saliences, once overall convergence has been obtained. Supplementary to the representation of the objects, the categories of each variable can be displayed graphically.

In addition to the above mentioned features, the new method allows for multiple points of view by clustering the categorical variables. Using the results from Meulman and Verboon (1993), it can be shown that the points of view loss function (1) can be interpreted as a weighted Euclidean individual differences model with within-subject equality restrictions on the dimension weights, where individual sources may only use p_s of the Σp_s dimensions in the overall analysis. In the next section, the minimization of the loss function (1) will be considered by separating the three sets of components $\{\mathbf{Y}_1,...,\mathbf{Y}_M\}$, $\{\mathbf{X}_1,...,\mathbf{X}_r\}$, and $\mathbf{A} = \{a_{ms}\}$.

4. Optimization

New quantifications for the categorical variables are found by reducing the stress in loss function (3) with respect to $\{\mathbf{Y}_1,...,\mathbf{Y}_M\}$, with values for $\{\mathbf{X}_1,...,\mathbf{X}_r\}$ and $\mathbf{A}$ considered fixed:

$$\mathrm{STRESS}(\mathbf{Y}_1,...,\mathbf{Y}_M) = (2NM)^{-1} \sum_{s=1}^{r} \sum_{m \in J_s} \|a_{ms}\mathrm{D}(\mathbf{G}_m\mathbf{Y}_m) - \mathrm{D}(\mathbf{X}_s)\|^2. \tag{3}$$

Finding updates amounts to solving a constrained multidimensional scaling problem for each variable separately. The basic theory for solving such problems, using majorization combined with projection, is in De Leeuw and Heiser (1980) and has been applied to other minimization problems in the context of fitting distance models to multivariate data in Meulman (1986, 1992). Detailed information on how to obtain an update $\mathbf{Y}_M$ can be found in the Appendix.

Then, keeping values for $\{\mathbf{Y}_1,...,\mathbf{Y}_M\}$ and $\mathbf{A}$ fixed, the procedure for updating $\{\mathbf{X}_1,...,\mathbf{X}_r\}$ can been reduced to a series of r relatively simple MDS problems, one for each point of view separately. The loss function considering the s-th point of view is written as

$$\text{STRESS}(\mathbf{X}_s) = (2NM)^{-1} \sum_{m \in J_s} \|a_{ms}\text{D}(\mathbf{G}_m\mathbf{Y}_m) - \text{D}(\mathbf{X}_s)\|^2,$$

and if at the outset a composite proximity matrix Θ_s is defined as

$$\Theta_s = M_s^{-1} \sum_{m \in J_s} a_{ms}\text{D}(\mathbf{G}_m\mathbf{Y}_m),$$

where M_S denotes the number of variables in subset s, then an interesting decomposition of STRESS($\mathbf{X}_s$), analogous to Heiser and De Leeuw (1977), can be obtained:

$$\text{STRESS}(\mathbf{X}_s) = (2NM)^{-1} \Big[\sum_{m \in J_s} \|a_{ms}\text{D}(\mathbf{G}_m\mathbf{Y}_m) - \Theta_s\|^2 + M_s \|\Theta_s - \text{D}(\mathbf{X}_s)\|^2 \Big]. \tag{4}$$

The first term on the right-side of (4) indicates how homogeneous the variables in subset s are with respect to the composite matrix Θ_s, and the second term, which is the only one dependent on $\mathbf{X}_s$, gives the badness-of-fit for subset s with respect to Θ_s. So the STRESS in (4) will decrease when the majorization approach to MDS is applied straightforwardly; further details are given in the Appendix.

The assignment of the M variables to subsets is performed at each iteration of the general algorithmic scheme. Keeping the values for $\mathbf{Y} = \{\mathbf{Y}_1,...,\mathbf{Y}_M\}$ and $\mathbf{X} = \{\mathbf{X}_1,...,\mathbf{X}_r\}$ fixed, a least squares loss function is considered as a function of $\mathbf{A}$ only, and minimized over all $\mathbf{A}$ satisfying $\mathbf{A} = \mathbf{WB}$. First, observe that because $\mathbf{X}_s$ is centered, $\mathbf{1}'\mathbf{X}_s = \mathbf{0}$, and normalized, $\|\mathbf{X}_s\|^2 = 1$, it follows that $\|\text{D}(\mathbf{X}_s)\|^2 = 2N$, and

$$\begin{aligned} \text{STRESS}(\mathbf{A}) &= (2NM)^{-1} \sum_{s=1}^{r} \sum_{m \in J_s} \|a_{ms}\text{D}(\mathbf{G}_m\mathbf{Y}_m) - \text{D}(\mathbf{X}_s)\|^2 = \\ &= (2NM)^{-1} \sum_{s=1}^{r} \sum_{m=1}^{M} \|a_{ms}\text{D}(\mathbf{G}_m\mathbf{Y}_m) - \text{D}(\mathbf{X}_s)\|^2 - (r-1). \end{aligned} \tag{5}$$

If $\tilde{\mathbf{A}}$ denotes an unconstrained solution to the function that considers all variables in all points of view, STRESS($\mathbf{A}$) = STRESS($\tilde{\mathbf{A}}$) + $M^{-1} \|\tilde{\mathbf{A}} - \mathbf{WB}\|^2$. Therefore (5) is initially minimized over elements in $\mathbf{A}$ unrestricted.

Since $\mathbf{G}_m\mathbf{Y}_m$ is centered and normalized, $\|\mathbf{G}_m\mathbf{Y}_m\|^2 = 1$ implies $\|\text{D}(\mathbf{G}_m\mathbf{Y}_m)\|^2 = 2N$, and an element $\tilde{a}_{ms}$ is then found, setting the partial derivatives with respect to a_{ms} in (5) equal to zero, as

$$\tilde{a}_{ms} = (2N)^{-1} \text{ tr } (\text{D}(\mathbf{G}_m\mathbf{Y}_m)'\text{D}(\mathbf{X}_s)).$$

As was mentioned in section 2, the estimate $\tilde{a}_{ms}$ reduces to Tucker's congruence coefficient due to the particular choice of normalization of $\mathbf{Y}_m$ and $\mathbf{X}_s$. So, for each variable, r congruence coefficients are obtained, one for each point of view $\mathbf{X}_s$, collected in $\tilde{\mathbf{A}} = \{\tilde{a}_{ms}\}$.

Because each variable is assumed to belong to one and only subset, the constrained $\mathbf{A}$ was defined as $\mathbf{A} = \mathbf{WB}$. The minimization of $\|\tilde{\mathbf{A}} - \mathbf{WB}\|^2$ over $\mathbf{W}$ and $\mathbf{B}$ can again be done for each variable consecutively. If $\max(\tilde{a}_{m1},...,\tilde{a}_{mr})$ denotes the largest element in the m-th row of $\tilde{\mathbf{A}}$, the m-th diagonal element w_{mm} of the diagonal matrix $\mathbf{W}$ is found as $\max(\tilde{a}_{m1},...,\tilde{a}_{mr})$. Next, the indicator matrix $\mathbf{B}$ is formed: b_{ms} is set to 1 if $\tilde{a}_{ms} = w_{mm}$, and 0 otherwise. It is unlikely that the procedure will result in a column s of $\mathbf{A}$ containing only zero's, i.e. none of the variables is assigned to the s-th point of view. But such a partition would in principle have no effect on the behavior of the algorithm, since the objective function is defined on proximities derived from variables belonging to a particular point of view.

Remark: The solution found for $\mathbf{A}$ is the global optimum for fixed $\mathbf{X}$ and $\mathbf{Y}$; however, in the overall algorithm, optimization is over $\mathbf{X}$ and $\mathbf{Y}$ as well, and local minima may occur, depending on how the process of assigning the variables to subsets is initialized. The algorithm developed for this paper uses the following procedure, explained for $r = 2$ but easily generalized to more points of view.

We start the algorithm by setting the number of points of view equal to 1, giving initial quantifications $\mathbf{Y}_m$, an initial weight vector $\mathbf{a}$, and a single space $\mathbf{X}$. Then, an $M \times M$ matrix Δ is formed with elements $\delta_{km} = 2N^{-1}\,\mathrm{tr}\,(\mathrm{D}(\mathbf{G}_k\mathbf{Y}_k)'\mathrm{D}(\mathbf{G}_m\mathbf{Y}_m))$; thus δ_{km} is the congruence coefficient between variable k and variable m. Next, a clustering algorithm, due to Hubert (1973), is applied to Δ to obtain an initial division of the index set $\{1,...,M\}$ into two subsets J_1 and J_2; this procedure can be briefly summarized as follows. First, order the elements of $\Delta = \{\delta_{km}\}$ for each pair of variables for which $k < m$. Select the pair having the smallest congruence, and assign these variables to different subsets. Then, inspect the pair of variables having the next smallest congruence. If $k \in \mathrm{J}_1$ or J_2, assign m to the other set (if $m \in \mathrm{J}_1$ or J_2, assign k to the other set). If none of the variables in the pair has been distributed as yet, do the assignment as soon as additional information has become available through a subsequent pair of variables. If both variables are already distributed, go to the next pair; proceed until all variables have been assigned.

After finding J_1 and J_2, we set $\mathbf{X}_s^o = \mathbf{X}$, $s = 1, 2$, and the elements of the columns in **A** are determined: if $m \in J_1$, $a_{m1} = a_m$, else $a_{m1} = 0$; if $m \in J_2$, $a_{m2} = a_m$, else $a_{m2} = 0$. On the basis of the initial estimates, new estimates for $\mathbf{X}_s$, **A**, and $\mathbf{Y}_m$ are computed consecutively, and iteration continues until the chosen stop criterion reaches some preset small value. At that point the partitioning algorithm is used a second time, to check whether its results are identical to the partitioning found by PVA. If they are not, iteration is continued, with the elements of the group structure matrix **B** changed according to the partitioning algorithm results. This approach has proven to work very well in practice, as will be illustrated in the next section.

5. Gauging Points of View Analysis

To study the behavior of the method, a data matrix **Z** was analyzed that has a particular 4-dimensional structure. The number of observations was set to $N = 100$, and the number of variables to $M = 12$. First, two sets of independent random samples $\mathbf{P} = \{\mathbf{p}_1,..., \mathbf{p}_8\}$ and $\mathbf{Q} = \{\mathbf{q}_1,..., \mathbf{q}_8\}$, of size N, were taken from two multivariate normal distributions $N(0, I)$. Then **P** was partitioned into $\mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2\}$, with $\mathbf{P}_1 = \{\mathbf{p}_1,..., \mathbf{p}_6\}$ and $\mathbf{P}_2 = \{\mathbf{p}_7, \mathbf{p}_8\}$, and **Q** into $\mathbf{Q} = \{\mathbf{Q}_1, \mathbf{Q}_2\}$, with $\mathbf{Q}_1 = \{\mathbf{q}_1,..., \mathbf{q}_6\}$ and $\mathbf{Q}_2 = \{\mathbf{q}_7, \mathbf{q}_8\}$.

Next, a 6×2 weight matrix **C** was constructed, whose rows are coordinates of points equally spaced on a circle with radius $(1-\alpha)^{1/2}$, where α denotes an a priori chosen amount of error, here set to $\alpha = 0.10$. Using **C**, two matrices **T** and **V** were created; the columns in **T** are formed by the following linear combinations: $\mathbf{t}_1 = \alpha^{1/2}(\mathbf{p}_1) + c_{11}(\mathbf{p}_7) + c_{12}(\mathbf{p}_8),...,\mathbf{t}_6 = \alpha^{1/2}(\mathbf{p}_6) + c_{61}(\mathbf{p}_7) + c_{62}(\mathbf{p}_8)$; in matrix formulation: $\mathbf{T} = \alpha^{1/2}\mathbf{P}_1 + \mathbf{P}_2\mathbf{C}'$. So **T** is of full rank, but because the two variables in $\mathbf{P}_2$ are included in each linear compound, **T'T** has two dominant eigenvalues. The matrix **V** was constructed analogously, using $\mathbf{Q}_1$ and $\mathbf{Q}_2$: $\mathbf{V} = \alpha^{1/2}\mathbf{Q}_1 + \mathbf{Q}_2\mathbf{C}'$. Finally, the continuous variables in **T** and **V** were transformed into categorical variables, with 5 categories each, by using optimal discretization (Max 1960), and the result was collected in the matrix **Z**, with $N = 100$ rows and $M = 12$ columns. For categorical data thus generated, it is expected that the category quantifications recover the original order of the categories, in analogy with the one-dimensional data structures studied in Meulman (1982), Van Rijckevorsel, Bettonvil, and De Leeuw (1985), and Gifi (1990).

To obtain some preliminary insight in the data structure to be analyzed, an ordinary linear, principal components analysis was performed first on the basis of **Z**, with the variables in deviations from their means and normalized with sums of squares equal to 1.0. The correlations of the original categorical variables with the first four components are given in Table 1; the component loadings themselves do not give a straightforward understanding of what is going on in the data. More can be seen from the representation of the loadings as points (in two pairs of dimensions in Figure 1). The ellipses that illustrate the structure among the variables were drawn in the plot of Dimension 2 versus 1, and 4 versus 3, but each other pair of dimensions would give elliptical structures as well. In the graph on the left, the variables 1–6 are on the dominant ellipse; in the graph on the right, this role has been taken by the variables 7–12.

Table 1. Principal Components Analysis in 4 dimensions: Component Loadings.

Variable	DIM1	DIM2	DIM3	DIM4
1	0.757	–0.269	–0.169	–0.483
2	0.795	0.239	–0.464	0.058
3	0.156	0.585	–0.307	0.647
4	–0.783	0.256	0.095	0.450
5	–0.802	–0.238	0.392	–0.137
6	–0.066	–0.564	0.382	–0.652
7	0.292	–0.635	0.357	0.508
8	0.374	0.165	0.821	0.220
9	0.045	0.760	0.456	–0.318
10	–0.343	0.586	–0.367	–0.527
11	–0.404	–0.172	–0.809	–0.173
12	–0.155	–0.740	–0.474	0.311

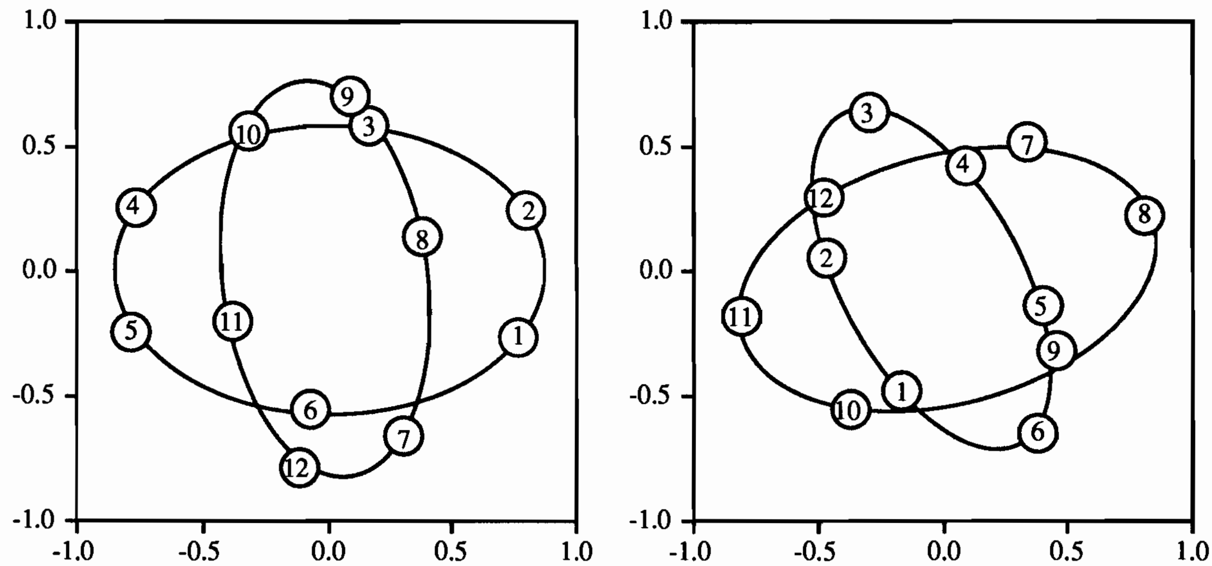

Figure 1. Loadings for variables from a 4-Dimensional principal components analysis of constructed data depicted as points. Dimension 2 versus 1 on the left; Dimension 4 versus 3 on the right. Ellipses connect variables that belong to the same subset according to how the data were constructed.

Results of a standard homogeneity analysis. Because the method proposed can be considered as an alternative to a standard homogeneity analysis, a HOMALS analysis (Gifi 1990) was performed first, with p = 4 dimensions in **X**. The fit of the variables in each HOMALS dimension are given by the so-called discrimination measures: $\|\mathbf{G}_m \mathbf{y}_{m(t)}\|^2$, $t = 1,\dots,p$, where $\mathbf{y}_{m(t)}$ denotes the t-th set of category quantifications for variable m; they are given in Table 2.

Table 2. Homogeneity Analysis in 4 dimensions: Discrimination Measures of Variables.
(* Denotes largest discrimination measure across dimensions.)

	HOM1	HOM2	HOM3	HOM4
1	0.356*	0.270	0.121	0.175
2	0.673*	0.108	0.123	0.024
3	0.216	0.124	0.406*	0.152
4	0.391*	0.068	0.100	0.295
5	0.686*	0.075	0.084	0.122
6	0.221	0.055	0.474*	0.200
7	0.242	0.072	0.553*	0.276
8	0.096	0.709*	0.183	0.363
9	0.310	0.495*	0.325	0.225
10	0.208	0.153	0.570*	0.294
11	0.023	0.703*	0.140	0.382
12	0.251	0.538*	0.125	0.323

If the variables are partitioned according to their largest discrimination measure across dimensions, the following tripartition would be obtained: {1, 2, 4, 5}, {8, 9, 11, 12}, {3, 6, 7, 10}. At this point, it is worthwhile to consider the 4 different transformations of the variables; these are depicted in Figure 2 (Dimensions 1 and 2) and Figure 3 (Dimensions 3 and 4). It turns out that the original order of the categories is perfectly recovered (with a slight exception for the second category of variable 4) in the dimension that has the largest discrimination measure for the variable. The sets {1, 2}, {8, 9} and {3, 10} show monotonic increase in Dimension 1, 2, and 3, respectively; because there were no order restrictions on the category quantifications, the sets of variables {4, 5}, {11, 12}, and {6, 7} are allowed to show monotonic decrease in these dimensions.

The representation of the object points is given in Figure 4 for Dimensions 1 and 2 (upper plot), and Dimensions 3 and 4 (lower plot). For the representation of a variable, we computed the correlations between the selected dimensions in **X** and one particular transformation: we used only quantifications if the discrimination measure in one of the two depicted dimensions were largest across all four dimensions. This criterion discarded subset {3, 6, 7, 10} from the upper plot in Figure 4. Next, the particular set of quantifications was used for which the discrimination measure was maximum (resulting, as seen from the transformation plots, in the use of quantifications that maintained either the original or the reversed rank order). So, in the upper plot for dimensions 1 and 2, the first set of quantifications was used for the variables {1, 2, 4, 5} and the second set for {8, 9, 11, 12}, while the lower plot depicts the remaining variables, {3, 6, 7, 10}, for which the third quantification was used.

The object points in the upper plot in Figure 4 are labeled with the original category numbers of variable 12; because variable 12 obtained a monotonically decreasing transformation, the second dimension orders the objects from category 1 at the top to category 5 at the bottom. This result can also been seen from the centroids of the objects according to this variable, which are also given. According to variable 11, the object points would be labeled in more or less the same way, while variables 8 and 9 would label objects at the top as 5's and objects at the bottom as 1's (because these variables have a monotonically increasing transformation).

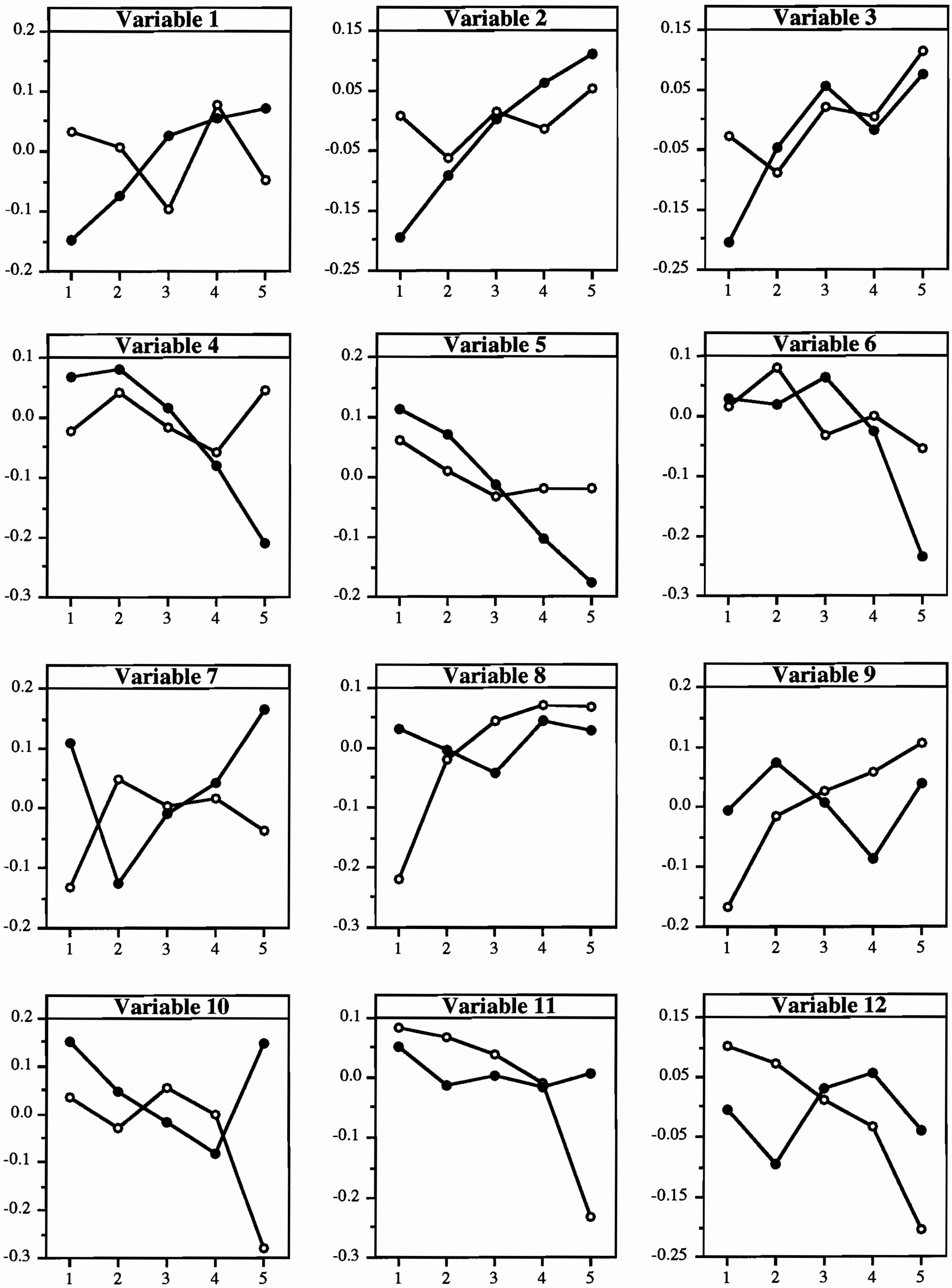

Figure 2.
Optimal quantifications in first two dimensions of standard homogeneity analysis of constructed data. The horizontal axis displays the original category numbers and the vertical axis the quantifications in dimension 1 (●) and 2 (○) simultaneously.

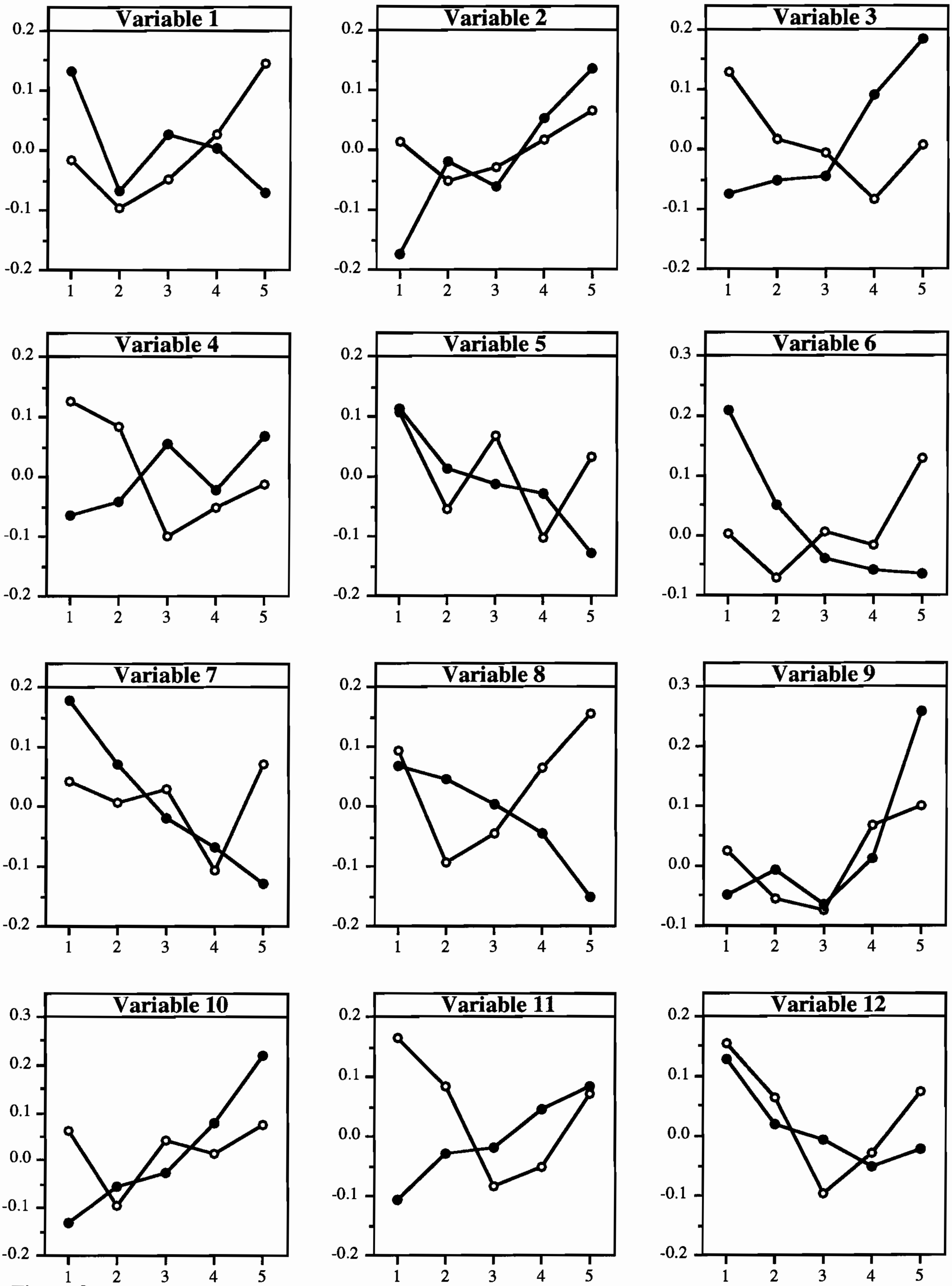

Figure 3.
Optimal quantifications in first two dimensions of standard homogeneity analysis of constructed data. The horizontal axis displays the original category numbers and the vertical axis the quantifications in dimension 3 (●) and 4 (○) simultaneously.

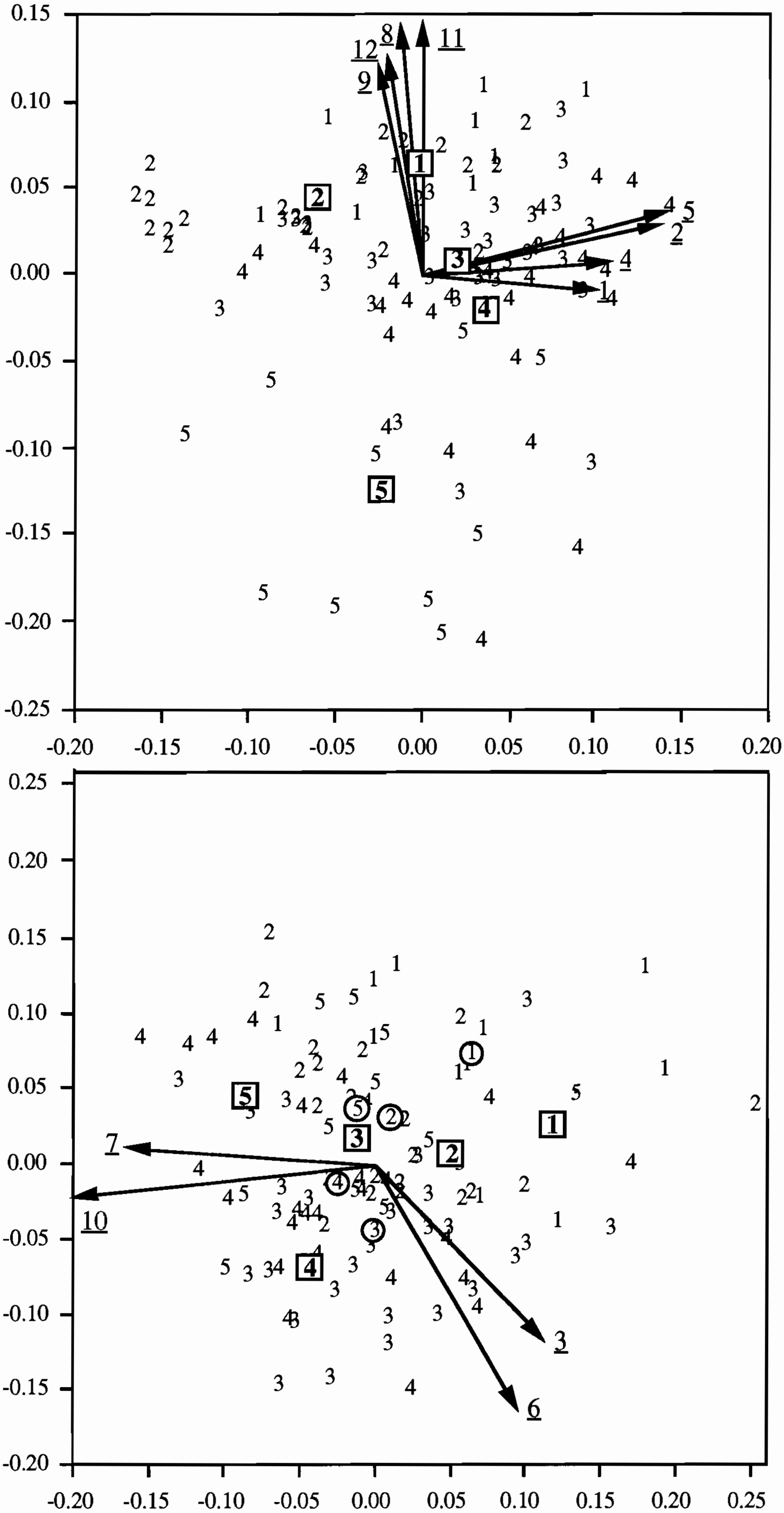

Figure 4. Object points and quantified variables in standard homogeneity analysis.

If the labeling had been done according to either variable 1 or 2, the horizontal axis would display 1's at the left and 5's at the right (and the reverse would be true for variable 5). Because there were no order restrictions, the analysis groups the variables: {1, 4}, {2, 5}, {8, 11} and {9, 12}; there is, however, not a very clear separation among the first two pairs, and hardly any among the last two pairs. Moreover, the variables {8, 9, 10 and 11} belong to the other subset, consistent to the manner in which the data were generated.

To show the contrast between the upper and lower plots in Figure 4, the points in the lower plot are labeled according to variable 12 as well; in addition, the centroids of the objects that belong in the same category according to variable 12 are displayed. The points of these (minor) centroids, denoted by circles, still show some dispersion in the space but not as much as a set of dominant centroids, which are denoted by a square and belong to variable 7. Because the third eigenvalue (0.267) is not much smaller than the second one (0.281), the HOMALS solution could also have been represented by displaying Dimension 3 versus 1 and 4 versus 2. In dimension 3 versus 1, the full first subset {1, 2, 3, 4, 5, 6} would have been displayed; but the variables 7 and 10 would have been represented in this space as well, because they have the largest discrimination measures in the third dimension. The overall conclusion, therefore, is that the HOMALS solution does not recover the original structure among the variables in a simple way. The major complication is that the analysis fails to separate the variables {3, 6} from {7, 10}.

Results of points of view analysis. The dimensionality of $\mathbf{X}_s$, $s = 1,...,r$, and $\mathbf{Y}_m$, $m = 1,...,M$, was set to 2, and the algorithm was started with $r = 1$. At the point where the matrix Δ with congruence coefficients between all pairs of proximity matrices was analyzed to obtain a partition into two subsets, the correct assignment {1, 2, 3, 4, 5, 6} and {7, 8, 9, 10, 11, 12} was recovered. Using this partition, iteration continued until convergence, giving final values for the weight matrix $\mathbf{A}$, the configurations $\mathbf{X}_1$ and $\mathbf{X}_2$, and the quantifications $\{\mathbf{Y}_1,...,\mathbf{Y}_{12}\}$; the assignment of the variables remained correct.

Table 3. Weights and Loss for Each Variable, and Partitioned Loss:
Mean Stress = Heterogeneity + Subset Stress = 1.0 - Mean Squared Weights

Variable	First Point of View Weight	Second Point of View Weight	Loss in Applicable Point of View Stress	Heterogeneity	Salience $G_m y_{m(1)}$ SSQ
1	0.899		0.193	0.144	0.786
2	0.900		0.190	0.142	0.808
3	0.866		0.251	0.169	0.748
4	0.889		0.210	0.151	0.760
5	0.897		0.196	0.143	0.800
6	0.874		0.236	0.169	0.743
7		0.867	0.248	0.165	0.748
8		0.886	0.216	0.153	0.785
9		0.898	0.193	0.147	0.730
10		0.876	0.233	0.163	0.750
11		0.887	0.213	0.152	0.781
12		0.890	0.208	0.151	0.772
		Mean Weight	Mean Stress	Hetero-geneity	Subset Stress
First point of view		0.887	0.106	0.076	0.030
Second point of view		0.884	0.109	0.078	0.031
Overall		0.886	0.215	0.154	0.061

With respect to the (rotated) quantifications, the first dimension was very dominant for all variables; the saliences are given in Table 3 that summarizes the various fit measures. The overall STRESS is 0.215 = 1.0 – 0.785, where the latter value is equal to mean of the squared elements in **A**. According to the different fit measures, the two points of view fit the two subsets of variables almost equally well. Figure 5 shows the two-dimensional quantifications for each variable in a single transformation plot. For subset #1, it is observed that the first quantification is monotonically increasing with the original category numbers for the variables 1, 2, and 3, while it is decreasing for 4, 5, and 6. For the second subset, monotonic increase is observed for variables 8, 9, and 10, and monotonic decrease for 7, 11, and 12. These patterns are the result from how $\mathbf{Y}_m$ was initialized (by taking averages of the associated object scores); whether the transformation shows monotonic increase or decrease has no influence, of course, on the derived distances. Some variables (2, 3, 5, 7, 8, 11) are almost perfectly linearly transformed (correlations with the original variable are larger than 0.985 or smaller than –0.987); in those cases, the second quantification shows a zigzag pattern. Departure from linearity in the first transformation also results in departure from the zigzag pattern in the second transformation; a first transformation with an S-shape (variable 9) results in a second quadratic transformation.

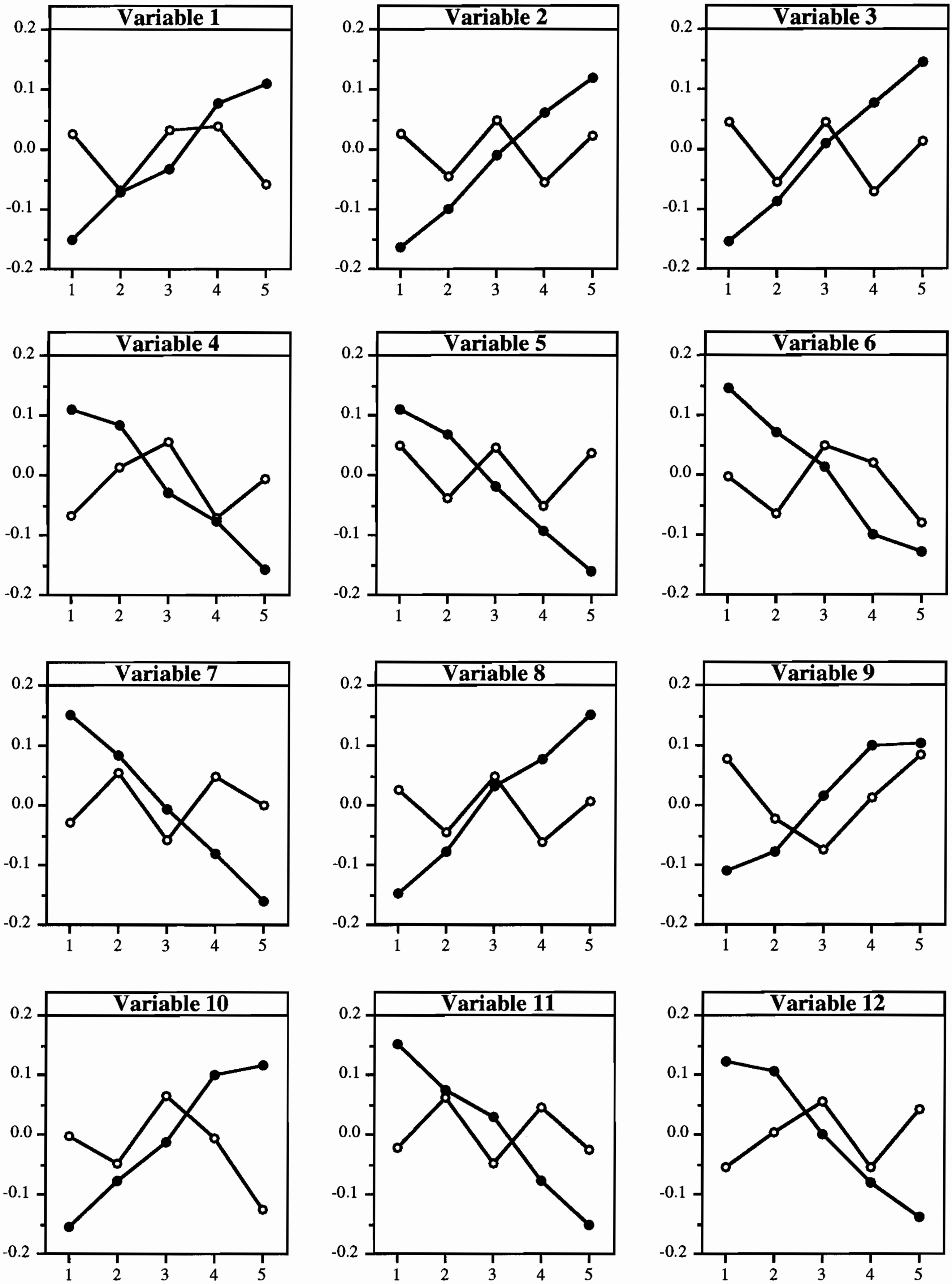

Figure 5.
Optimal quantifications in two dimensions of points of view analysis of constructed data. The horizontal axis displays the original category numbers and the vertical axis the quantifications in dimension 1 (●) and 2 (○) simultaneously.

Figure 6 displays the objects in $\mathbf{X}_1$ (upper plot) and $\mathbf{X}_2$ (lower plot); the points are labeled with the original category numbers of variable 2. It is clear that the horizontal axis in the upper plot orders the objects from category 1 at the left up to category 5 at the right. As a diagnostic, the centroids of objects in the same category are displayed as well. The figure also gives vectors for the six variables belonging to the first subset. The coordinates for the m-th variable were obtained from the correlations between columns of $\mathbf{X}_1$ and $\mathbf{G}_m\mathbf{y}_{m(1)}$, where $\mathbf{y}_{m(1)}$ denotes the first set of category quantifications for variable m. Because the transformation function for the variables 4, 5, and 6 was allowed to be decreasing with the original category numbers, three groups of variables are obtained: {1, 4}, {2, 5}, and {3, 6}, positioned almost perfectly on a semicircle with (almost) equal spacing, which is completely consistent with how the variables were generated.

The lower plot in Figure 6 gives the object points in $\mathbf{X}_2$, which models distances for the variables {7, 8, 9, 10, 11, 12}; to illustrate that this space is dramatically different from $\mathbf{X}_1$, the points are labeled again according to variable 2 (that belongs to the first subset). Also, the centroids of the objects that belong in the same category according to variable 2 are displayed; the points for these (minor) centroids, denoted by circles, are grouped very closely around the origin of the space, which indicates that the quantified variable 2 does not fit in $\mathbf{X}_2$. To demonstrate the contrast, the centroids in agreement with variable 12 are depicted as well by squares. In space $\mathbf{X}_2$, the same phenomenon occurs as was observed in $\mathbf{X}_1$: because of either increasing or decreasing transformations, the variables group in pairs, positioned with ordering {11, 8}, {12, 9}, {7, 10} on a semicircle. Again, the spacing between the groups of variables is close to being equal.

The correlations of the 2×2 dimensions in the points of view analysis and the four dimensions in the HOMALS analysis show that the first dimension in the first point of view has the largest correlation (0.851) with the first HOMALS dimension. The second dimension in the first point of view correlates maximally (0.562) with the third HOMALS dimension. The first dimension in the second point of view has the largest correlation (0.791) with the second HOMALS dimension, and the second dimension in the second point of view with the third HOMALS dimension (0.770).

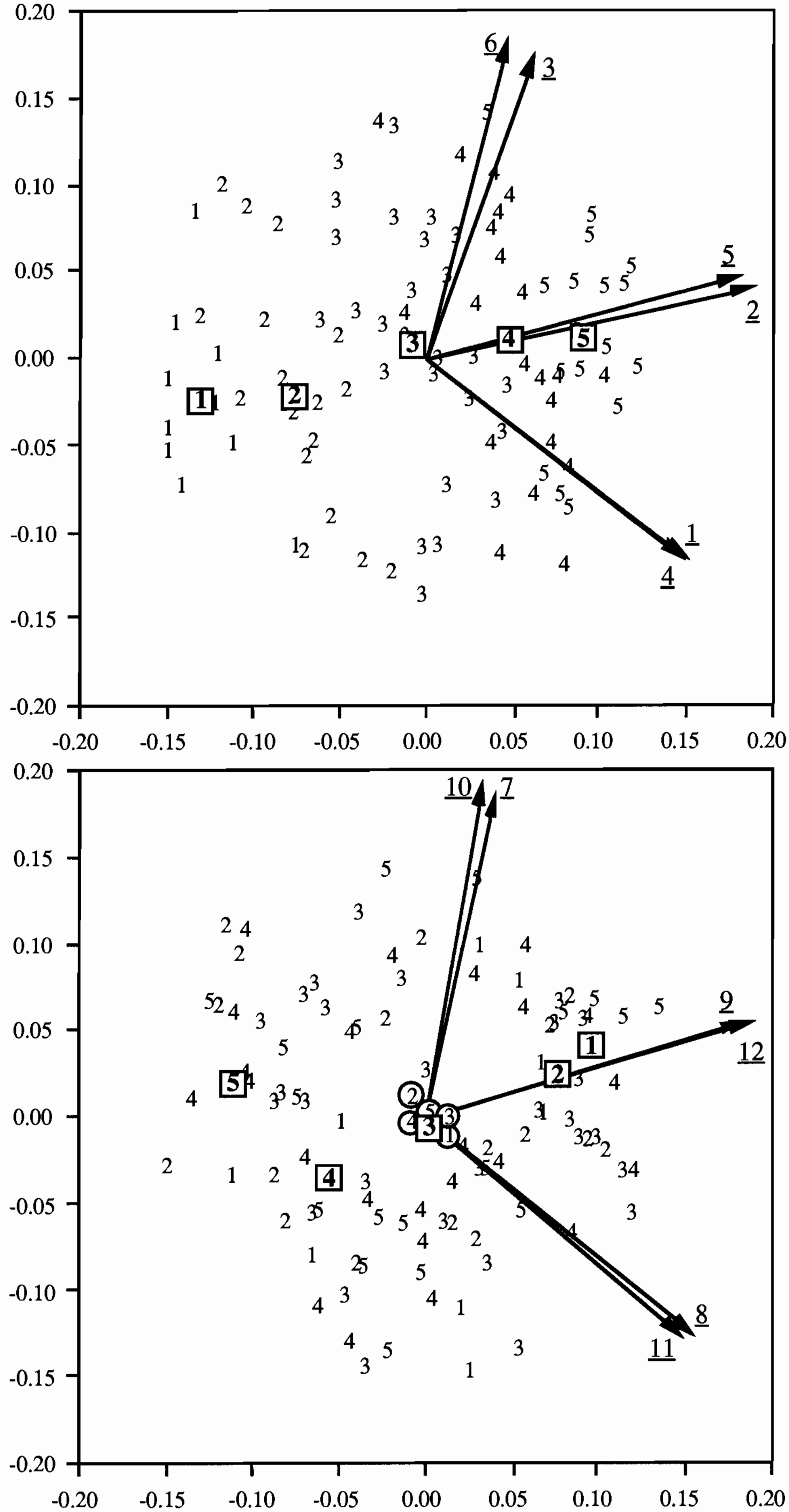

Figure 6. Object points and quantified variables in points of view analysis.

The points of view analysis was done with two dimensions in $\mathbf{Y}_m$, and four dimensions in $\mathbf{X} = \{\mathbf{X}_1; \mathbf{X}_2\}$; in the standard HOMALS analysis the number of dimensions in $\mathbf{Y}_m$ is necessarily equal to the number of dimensions assumed in **X**. Moreover, because PVA is defined on distances, we had the freedom to rotate the quantifications in $\mathbf{Y}_m$, by which the original rank orders were recovered in the first set of quantifications. The HOMALS loss function is defined on the inner products of **X** and the $\mathbf{Y}_m$, and so there is no freedom to rotate each $\mathbf{Y}_m$ separately. The original rank orders were only recovered by inspecting the discrimination measures, and they appeared in different dimensions for different sets of variables. The points of view analysis separated the two subsets of variables perfectly; the HOMALS analysis failed to do so. As remarked in Section 3, however, a drawback of PVA compared to HOMALS is that the centroid principle no longer applies.

Permutation Tests. Using quadratic assignment measures of concordance, permutation tests (Hubert 1987) are available to evaluate the results of the points of view analysis. Adopting the congruence coefficient, one can evaluate whether multiple points of view are truly different and also whether a variable has no communality with the subset to which it is not assigned. Applying permutation tests in the current application gives the following results.

The observed congruence, $c_{(o)}$, between $D(\mathbf{X}_1)$ and $D(\mathbf{X}_2)$ is 0.815 (which corresponds to a Pearson correlation, based on vectors created from the elements in the lower-triangular matrices, of –0.002). By computing the congruence coefficient, $c_{(p)}$, in a large number permutations, say 1000, of the rows and columns in $D(\mathbf{X}_2)$, the probability of a congruence coefficient as large or larger than the observed value, $P(c_{(p)} \geq c_{(o)})$, occurring by a random displacement of the points in $\mathbf{X}_2$ can be estimated. The results are given in Table 4; because the *P*-value is 0.54, there is no evidence of communality between $\mathbf{X}_1$ and $\mathbf{X}_2$. For each variable separately, the observed congruence between the derived proximities and the distances in the space fitted to the subset to which the variable has *not* been assigned, can be evaluated against the congruence obtained by random displacement of the points in this space. For all variables (Table 4), the *P*-value exceeds the conventional value of 0.05; therefore it may be concluded that there is no evidence of communality between variables and the subset to which they were not assigned.

Table 4. Permutation Test for Congruence Coefficient for the two Points of View and with Respect to the Point of View to which a Variable Was *Not* Assigned. 1000 Permutations to Estimate $P(c_{(p)} \geq c_{(o)})$.

For **X**: $c_{(o)} = (2N)^{-1}$ tr $(D(\mathbf{X}_1)'D(\mathbf{X}_2))$;
$c_{(p)} = (2N)^{-1}$ tr $(D(\mathbf{X}_1)'D(\mathbf{X}_{2(p)}))$.

For m = 1,...,6: $c_{(o)} = (2N)^{-1}$ tr $(D(\mathbf{G}_m\mathbf{Y}_m)'D(\mathbf{X}_2))$;
$c_{(p)} = (2N)^{-1}$ tr $(D(\mathbf{G}_m\mathbf{Y}_m)'D(\mathbf{X}_{2(p)}))$.

For m = 7,...,12: $c_{(o)} = (2N)^{-1}$ tr $(D(\mathbf{G}_m\mathbf{Y}_m)'D(\mathbf{X}_1))$;
$c_{(p)} = (2N)^{-1}$ tr $(D(\mathbf{G}_m\mathbf{Y}_m)'D(\mathbf{X}_{1(p)}))$.

	Observed	Maximum	Mean	Minimum	*P*-value
X	0.815	0.834	0.816	0.798	0.54
1	0.760	0.780	0.762	0.749	0.67
2	0.760	0.783	0.761	0.744	0.51
3	0.751	0.774	0.751	0.732	0.48
4	0.753	0.778	0.758	0.743	0.78
5	0.759	0.783	0.758	0.743	0.40
6	0.757	0.783	0.754	0.737	0.33
7	0.750	0.771	0.753	0.736	0.70
8	0.747	0.779	0.754	0.735	0.87
9	0.763	0.779	0.759	0.746	0.17
10	0.750	0.780	0.750	0.733	0.50
11	0.755	0.780	0.755	0.736	0.47
12	0.763	0.781	0.758	0.743	0.16

Incorrect initial assignment. To see how the procedure behaves given an incorrect initial assignment of variables to subsets, the algorithm was used again up to the point where the clustering algorithm had distributed {1, 2, 3, 4, 5, 6} to subset #1 and {7, 8, 9, 10, 11, 12} to subset #2. From here, a Monte Carlo study was performed, using random starts. The first important result was that local minima (incorrect final subsets) were found especially because the 12 variables in the constructed data consist of 6 paired variables (these are variables that have (almost) perfect negative correlation in the original data, such as {1, 4}, {2, 5}, {3, 6}, etc.). Whenever the random start assigned paired variables to the incorrect subset, the PVA algorithm managed to move other variables to the correct subset, but failed to move the particular pair(s). To separate this very special property of the data from the remaining local minima problem, the Monte Carlo study was continued by excluding initial starts that assigned paired variables to the wrong subset. This approach automatically excluded the undesirable partitions with one versus 11 and two versus 10 variables, from the initial starts; these are undesirable because the dimensionality p_s was set to 2.

Each other combination that creates two subsets out of twelve variables was given an equal probability; the number of samples taken was 100. The algorithm was stopped when it had recovered the correct subsets; when it seemed to fail to find the correct partition, a stop criterion (.0001) was used, defined on the largest change occurring in $\tilde{\mathbf{A}}$. When the algorithm had stopped initially (either with the correct or the incorrect partition), the clustering algorithm was applied to give possibly different assignments, and iteration was resumed. The results are given in Table 5 and can be summarized as follows.

Table 5. Results of the Monte Carlo Study with Random Starting Points.

		Percentage of Correct Assignments for Individual Variables	
Solutions	N	Initially	Finally
Total	100	97	100
Correct	75	100	100
Incorrect	25	89	100
Variable	1	99	100
Variable	2	99	100
Variable	3	97	100
Variable	4	100	100
Variable	5	100	100
Variable	6	76	100
Variable	7	100	100
Variable	8	99	100
Variable	9	99	100
Variable	10	99	100
Variable	11	99	100
Variable	12	100	100

From the 100 starting points, 99 of which were incorrect, the correct partition was found initially in 75% of the cases. Apart from this global result, partitions were inspected on the level of individual variables. In the incorrect initial partitions, an average of 89% of the variables was assigned correctly, which gives an overall 97% initial correct assignment of the variables. Looking at the variables separately, all are initially assigned correctly in at least 97% of the cases, except for variable 6 (76%), so the latter variable is assigned incorrectly in 24% of the cases and is therefore responsible for almost all of the 25% of incorrect partitions. When the PVA algorithm had initially stopped, either with the correct or an incorrect partition, resuming iteration on the basis of the result of the clustering algorithm, using the current values of $D(\mathbf{G}_m\mathbf{Y}_m)$, provided the correct solution in 100% of the cases.

In Figure 7, the separate histories (in the first 50 iterations) of the congruence coefficients for variable 6 and 8 with each point of view (given in $\tilde{\mathbf{A}}$) are shown for a particular successful case, with starting point {1, 2, 3, 10, 11, 12} and {4, 5, 6, 7, 8, 9}. Variables 8 and 6 both start in subset 2; the paths of the congruence coefficients for variable 8 diverge after iteration #7, while paths for variable 6 cross at iteration 17.

The assignment process using the clustering algorithm was also examined as follows. First, 100 starting points were generated, each by applying the clustering algorithm to a matrix with congruence coefficients based on a random sample of size $N = 80$ from the 100 rows of the initial $\mathbf{G}_m\mathbf{Y}_m$. With a smaller sample size, the particular (2×2)-dimensional structure is made quite less prominent. Using these starting points, the PVA algorithm continued with the analysis of the full data set, with $N = 100$.

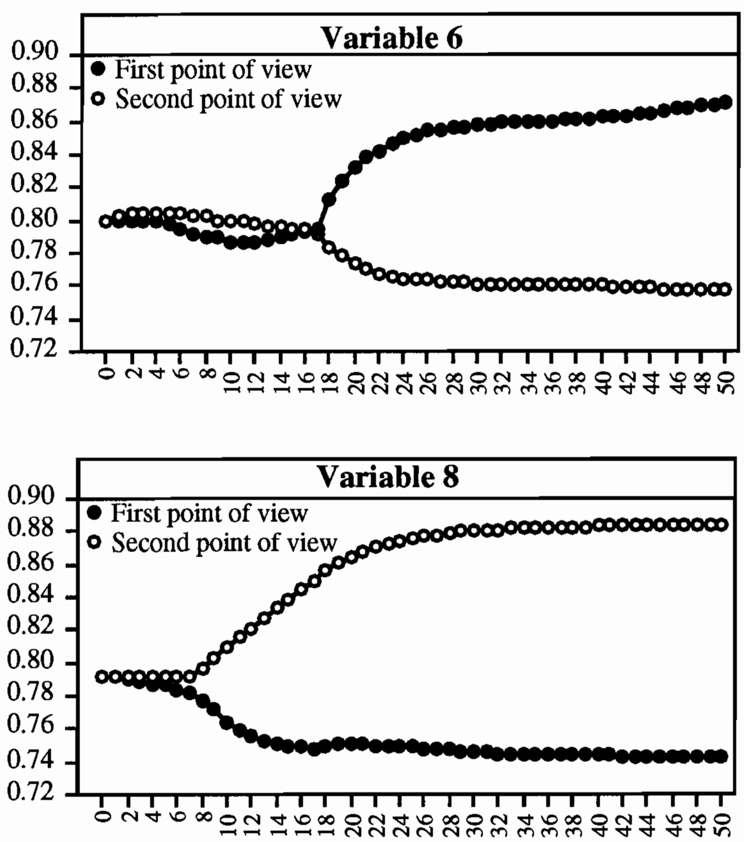

Figure 7. Two separate histories (in the first 50 iterations) of congruence coefficients for variable 6 and 8 with each point of view. Starting point: {1, 2, 3, 10, 11, 12} and {4, 5, 6, 7, 8, 9}. Paths of the congruence coefficients for variable 8 diverge after iteration #7; paths for variable 6 cross at iteration 17.

From the 100 samples taken, the clustering algorithm found the correct partition as a starting point in 74% of the cases. These samples also ended with the correct partition. Occurrence of paired variables assigned to the wrong subset was allowed this time. From the remaining 26 incorrect starting points, 16 did not involve paired variables in the wrong subsets; from these 16 starting points, 14 resulted in a correct initial partition. From the ten wrong partitions that involved paired variables, only one ended initially in the correct partition. In the 11 cases that the PVA algorithm did not seem to find the correct partition initially, subsequent use of the clustering algorithm and resuming iteration resulted in correct assignments for all cases. Therefore, the analysis ended correctly for 100% of the final solutions.

We conclude as a general analysis strategy that the use of the clustering algorithm should be definitely preferred over random starts. Also, it seems extremely worthwhile to apply the clustering algorithm again when the overall PVA algorithm has initially converged and to resume the iteration process when the clustering algorithm gives a partition different from the current one.

6. Application

The data analyzed in the application were kindly made available by Professor Louise Fitzgerald, Department of Psychology, University of Illinois at Urbana-Champaign. The items resulted from an extensive enumeration of the myths concerning rape, through examination of previous research and writing, the experiences of women victims, and similar sources. Following this enumeration, a pool of items in the form of attitudinal statements was developed, and observations on the items were collected from multiple samples of both sexes by the so-called free-sort method: the subjects were allowed to classify the statements (subjectively) into groups according to the similarity features they attributed to the statements, leaving the choice of number of clusters formed up to the subjects. The complete data consists of the results of the free-sort of samples from the item pool obtained for four groups of women and four groups of men. The points of view analysis was applied to a single sample, with 19 male and 18 female subjects sorting 40 statements. Each subject also attributed labels to the categories that he or she used to sort the statements; the particular set of statements is given in Table 6.

Table 6. Statements on Rape Myth given in the Free-Sort Task.

1	To some degree, women provoke rape by their appearance or behavior.
2	Given the way some women dress, you can't blame guys for trying to get as much as they can.
3	If a woman is hitchhiking, it is partly her own fault if she is raped.
4	If a woman is raped while she is drunk, she is at least somewhat responsible for letting things get out of control.
5	If a woman is out alone at night and she is raped, it's partly her own fault.
6	If a woman gets drunk at a party and has sex with a man she just met there, she can be considered "fair game" to other men at the party to have sex with her too.
7	Some women just deserve to be raped.
8	It would do some women good to be raped.
9	Some women just need some rough sex now and then to show them who's boss.
10	It is understandable that men sometimes force sex on women who tease them sexually.
11	Women who lead men on get what they deserve.
12	When a guy forces his date to have sex, it's generally only because he's really drunk.
13	Forcing a woman to have sex is forgivable if the man was drunk and just couldn't stop himself.
14	Women are generally attracted to sexually forceful men.
15	After a man forces himself on a woman sexually, she may start enjoying it.
16	Frequently, women who protest against having sex end up enjoying themselves.
17	If a woman is raped, it is probably because she didn't fight back enough.
18	A woman cannot really be "raped" against her will.
19	If the rapist doesn't have a weapon, it probably isn't a rape.
20	If a woman is not physically hurt, there is no way to tell if she was really raped.
21	When a man pays for an expensive date, the woman owes him some sexual activity.
22	If a man pays a lot of money on a date, she should have sex with him if he wants to.
23	If a man spends a lot of money on dinner and drinks, he is justified in expecting sex.
24	Women are often raped because they didn't say no clearly enough.
25	Rape isn't as big a problem as some people think.
26	If a woman is going to be raped, she might as well relax and enjoy it.
27	Good looking women are more likely to be raped than women who are physically unattractive.
28	Because men are physically stronger than women, rape will always happen.
29	For a man to force sex on a woman is only natural.
30	Rape is nature's way of insuring that only the most aggressive and fittest men reproduce.
31	Women are not generally raped by men they know.
32	Most rapists are seriously mentally ill.
33	Women who dress suggestively are more likely to be raped.
34	Lower class women are more likely to be raped than other women.
35	Prostitutes can't really be raped.
36	If a woman accuses her former lover of raping her, it probably isn't really true.
37	Women who are embarrassed by a casual sexual encounter may afterwards label it rape to cover for themselves.
38	Many so-called rape victims are actually women who had sex and "changed their minds" afterwards.
39	A rape charge days or months after the incident is supposed to have occurred is probably not really rape.
40	Women sometimes make up rape charges for no apparent reason.

Any multivariate analysis of these data should take the following into account. The items (the statements) are most appropriately considered as objects in the analysis, and the subjects as variables, since the subjects attributed structure to the items and not conversely. The items display a fundamentally unordered categorical structure that might differ from subject to subject. From the class of existing multivariate analysis techniques, a standard homogeneity analysis can be considered appropriate. HOMALS was applied with $p = 2$ dimensions in **X**, giving 0.772 as overall fit (the mean of the eigenvalues in two dimensions). The first two dimensions of the item space **X** are displayed in Figure 8, and the configuration typically shows a horse-shoe solution, also known as the Guttman effect: the second dimension is a quadratic function of the first one, and therefore the conclusion would be that the data contain a dominant one-dimensional structure. In addition, three items (21, 22, and 23) are distinctly separated from the remaining ones, and at the bottom of the configuration a considerable number of points are clustered together.

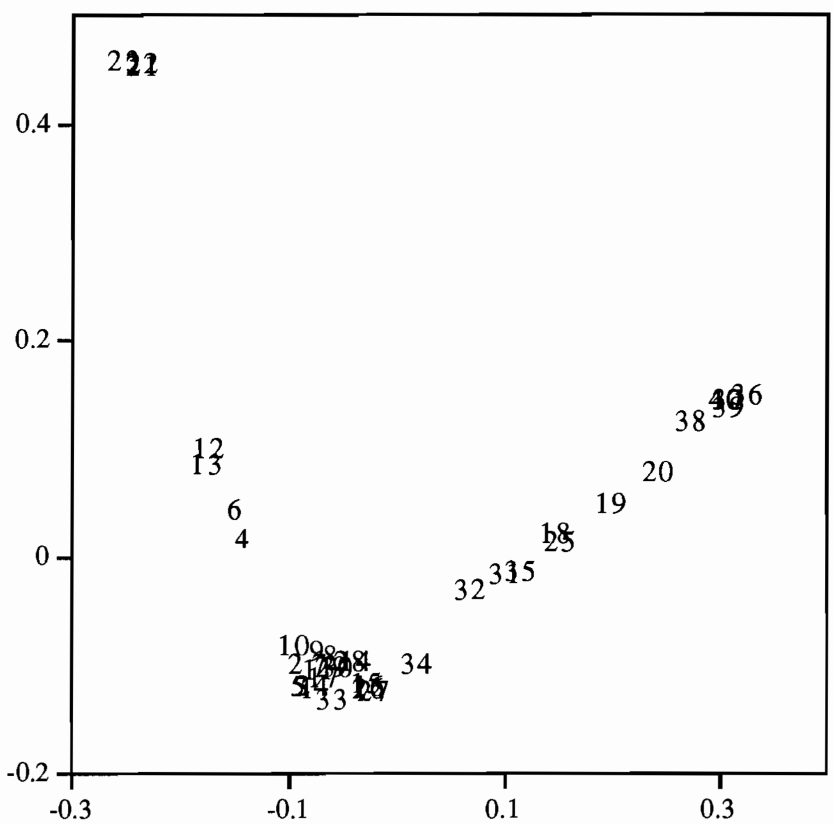

Figure 8. Rape myths items as represented by a two-dimensional homogeneity analysis. Second dimension (vertical axis) versus first dimension (horizontal axis).

As mentioned above, the proposed points of view analysis differs in at least two aspects from a standard homogeneity analysis. First, the space for the items is optimal with respect to distances derived from the quantifications, and second, a points of view analysis would show whether subjects use similar or different mechanisms in the allocation of items to particular groups. One obvious research question would be whether women perform the sorting task differently from men, which can be inspected through two separate analyses, or by using the present algorithm with **B** (the group structure matrix) fixed. The raw STRESS equals 325.84, associated with a mean squared congruence of 0.890, and the mean squared canonical correlation between the dimensions in the two spaces is 0.833. Thus, the badness-of-fit value is acceptable, and the female and male points of view are clearly very similar.

Next, the algorithm was used to find two groups empirically. The clustering algorithm separated subjects 13, 18, 25, 27, 31, 34, and 37 from the others; during the iterative process, subjects 13 and 18 were moved to the first cluster of variables. The raw STRESS obtained was 295.92, associated with a mean squared congruence of 0.900, and the mean squared canonical correlation between the dimensions in the two spaces is 0.376. In this solution, the STRESS is clearly less than in the female versus male analysis, and the two points of view are less similar. In the standard homogeneity analysis, a fit measure for the variables in dimension t is given by the discrimination measure: $\|\mathbf{G}_m\mathbf{y}_{m(t)}\|^2$. For the subjects 25, 27, 31, 34 and 37 who fit better in the second point of view, HOMALS gives a grand mean discrimination measure over two dimensions of 0.430, which is far below the average of 0.772. Figure 9 displays the values of the congruence coefficients between all variables and the two points of view at the point of convergence. The coefficients have been sorted with respect to the first point of view, and it is clear that the discrepancy between the values for the two different viewpoints is sufficiently large.

The two points of views are displayed in Figures 10 and 11. The points have been labeled with the original item number; the first point of view shows a number of clusters that have been given a substantive interpretation through the use of the labels subjects attributed to their classification.

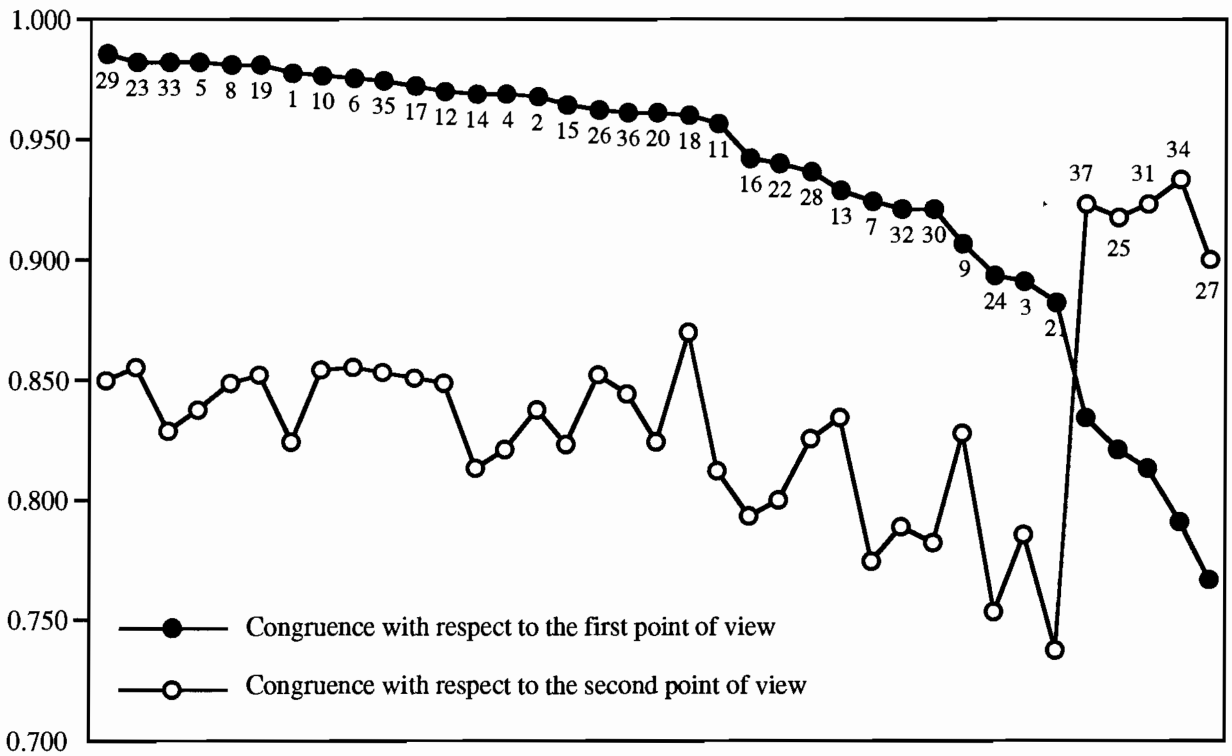

Figure 9. Congruence coefficients between rape myth item proximities and two points of view.

When inspecting the individual items, it is clear that the first HOMALS dimension is highly correlated with the first dimension in the first point of view; it accounts for 98% of the variance. The first dimension separates items concerning sex that is expected on dates (e.g., "When a man pays for an expensive date, the woman owes him some sexual activity") and rape related to the use of alcohol ("If a woman gets drunk at a party and has sex with a man she just met there, she can be considered 'fair game' to other men at the party to have sex with her too") from items that state that either the woman was lying ("Many so-called rape victims are actually women who had sex and 'changed their minds' afterwards") or that it could not have been rape ("A woman cannot really be 'raped' against her will"). In between these items, several clusters of items are found, separated in the second dimension, that were completely grouped together in the HOMALS solution. At the top, items are found that blame rape to appearance or behavior ("Given the way some women dress, you can't blame guys for trying to get as much as they can"), followed by items that say that either women deserve rape and/or describe rape as the result of nature itself ("Because men are physically stronger than women, rape will always happen"). At the bottom, a cluster of statements is found that relate rape to pleasure ("If a woman is going to be raped, she might as well relax and enjoy it"). If the second point of view is considered at the individual item level, we notice that particular items are grouped in small clusters that have very close resemblance to groupings in the first point of view. Using linear or quadratic assignment measures of concordance, permutation tests (Hubert 1987) are available to evaluate the similarity. Permutation tests, using a random relabeling mechanism 1000 times, gave p-values of 0.00, so the overall similarity as measured by the mean squared canonical correlation and the congruence coefficient is significant. However, given the clusters, proximity between clusters is displayed rather differently. Groups of items that were clearly separated from the rest, like items 21, 22, and 23 ("Sex is expected"), are now close to other clusters. Subjects in the second point of view do not seem to discriminate items as much as was done in the first point of view.

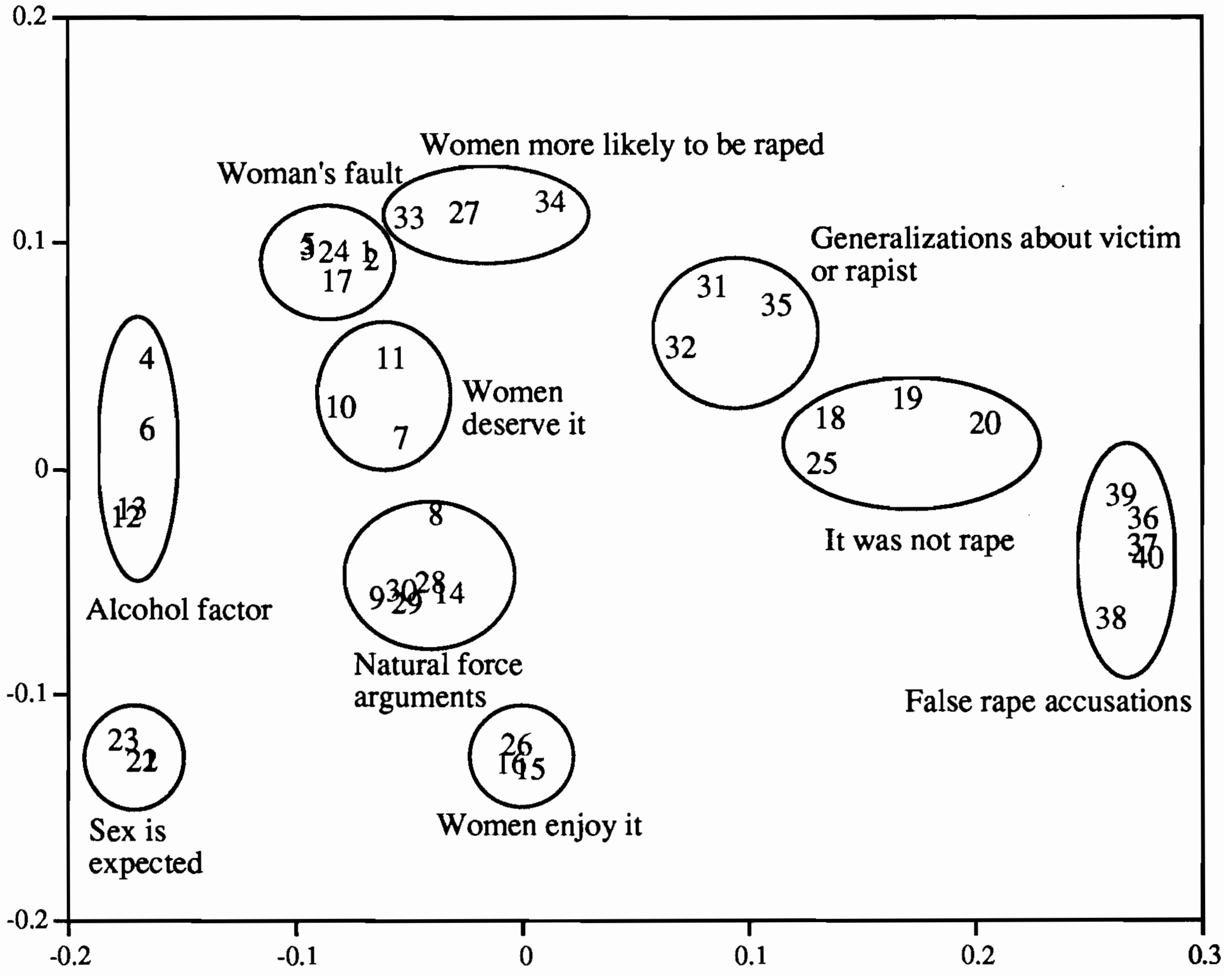

Figure 10. Rape myths items as represented by points of view analysis. Second dimension (vertical axis) versus first dimension (horizontal axis) of the first point of view.

Items at the left-side are all "reasons why it's the women's fault", while items at the right-side give reasons why "rape isn't that big of a problem". At the same time, subjects in the second point of view do not perceive particularly strong similarity among other items; for instance, items 36 up to 40, positioned in a tight cluster at the right-side of the first dimension in the first point of view, are scattered around in the second point of view (they have been connected by lines). The encircled items are the most influential ones, in the sense that their removal from the two configurations would increase the mean squared canonical correlation between the points of view. Given the positions of the items, relabeling of the points by a pairwise interchange procedure gives a mean squared canonical correlation as large as 0.797, which is considerably higher than the original value 0.376. After this relabeling, only 3 points had retained their original label.

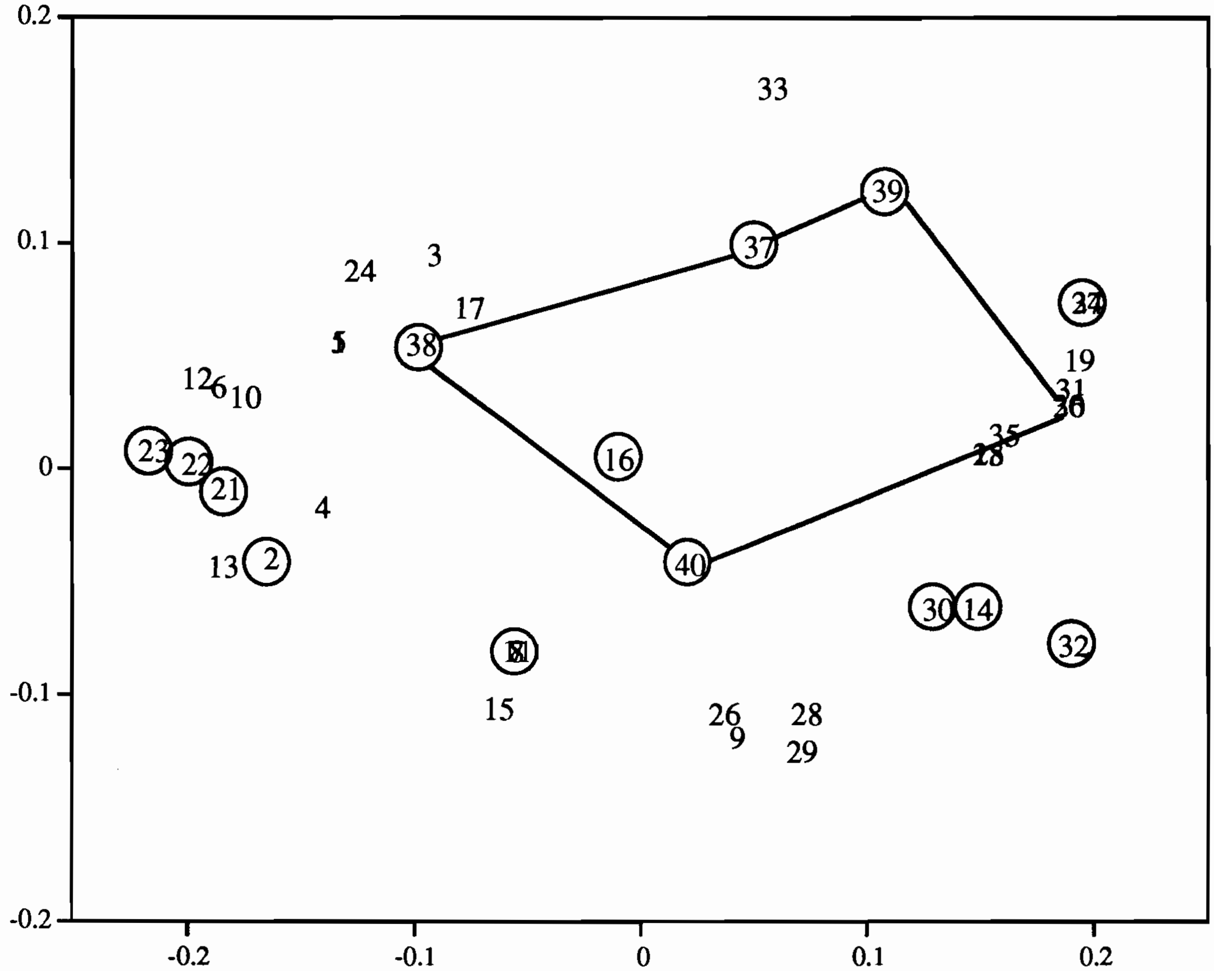

Figure 11. Rape myths items as represented by points of view analysis. Second dimension (vertical axis) versus first dimension (horizontal axis) of the second point of view.

Therefore, we conclude that although the two points of view are not totally different, the second point of view has picked up some deviant classification rationales from the five associated subjects, who are all male.

Since the Monte Carlo study presented above showed that the use of the Hubert (1983) algorithm very obviously outranks the use of random starts for the initialization of **B**, such a study will not be repeated here. Instead, to obtain some insight as to whether the fit measures obtained are different from those for other partitions with the same number of subjects, 100 solutions were computed with two points of view, always representing 32 and 5 subjects, respectively. Results with respect to STRESS give 340.24 as the maximum, 329.36 as the mean, and 300.34 as the minimum, so the partition found originally clearly has a better fit. The results for the 100 values for the mean squared canonical correlation are 0.994 (maximum), 0.712 (mean), and 0.408 (minimum), so the item spaces are more similar than in the original solution.

To conclude the analysis, a Jackknife study was done, leaving one subject out at a time. Given a grand mean of squared canonical correlations between the original solution and the 37 Jackknife solutions of 0.981 for the first point of view and 0.956 for the second point of view, the conclusion can be made that the two points of view are sufficiently stable.

7. Discussion

The procedure presented appears to be well suited for analyzing a heterogeneous set of categorical variables, assuming they can be partitioned into homogeneous subsets. Concerning the local minimum problem, we conclude that the use of random starts may give suboptimal solutions. However, even with random starts the supplementary use of the partitioning algorithm for evaluating the final solution and resuming iteration in the case of a different result, turned out to be effective in preventing the occurrence of local minima. When the partitioning algorithm is used at the outset instead of a random start, local minima appear unlikely to occur.

A major choice in the proposed procedure is to define the loss function to be minimized on distances. This choice can be motivated by giving major emphasis to the representation of the individual objects, so their mutual distances are optimally approximated. The application showed that in the case of a dominant first dimension, the use of least squares distance approximation gives a substantively different second dimension, whereas a standard homogeneity analysis gives a horseshoe, the second dimension being a quadratic function of the first one. The drawback of least squares distance approximation, of course, is that the computational burden may become too heavy when the number of objects becomes large. In such cases, however, our main interest will probably not be in the position of the individual objects, but in that of particular groups and the optimal quantification of the variables. If we still wish to cluster the variables into homogeneous subsets, a similar approach to that discussed in this paper may be developed for classical homogeneity analysis as well; research in this direction will be forthcoming.

Appendix

The algorithm applied the majorization approach to MDS at two different points. First, having the decomposition in (4), an update for $\mathbf{X}_s$ that reduces the value of $\|\Theta_s - \mathrm{D}(\mathbf{X}_s)\|^2$ is given from a starting point $\mathbf{X}_s^o$ by:

$$\mathbf{X}_s = N^{-1}\mathrm{B}(\mathbf{X}_s^o)\mathbf{X}_s^o.$$

Here, the $N \times N$ matrix $\mathrm{B}(\mathbf{X}_m^o)$ can be written as

$$\mathrm{B}(\mathbf{X}_m^o) = \mathrm{B}^*(\mathbf{X}_m^o) - \mathrm{B}^o(\mathbf{X}_m^o), \tag{6}$$

where the elements of the matrix $\mathrm{B}^o(\mathbf{X}_m^o)$ are given by

$$\mathrm{b}_{ij}^o(\mathbf{X}_s^o) = \frac{\theta_{ij(s)}}{\mathrm{d}_{ij}(\mathbf{X}_s^o)} \qquad \text{if } i \neq j \text{ and } \mathrm{d}_{ij}(\mathbf{X}_s^o) \neq 0;$$

$$\mathrm{b}_{ij}^o(\mathbf{X}_s^o) = 0 \qquad \text{if } \mathrm{d}_{ij}(\mathbf{X}_s^o) = 0,$$

with

$$\theta_{ij(s)} = M_s^{-1} \sum_{m \in \mathrm{J}_s} a_{ms}\mathrm{d}_{ij}(\mathbf{G}_m\mathbf{Y}_m),$$

and the elements of the diagonal matrix $\mathrm{B}^*(\mathbf{G}_m\mathbf{Y}_m^o)$ by

$$\mathrm{b}_{ii}^*(\mathbf{X}_s^o) = \mathbf{1}'\mathrm{B}^o(\mathbf{X}_s^o)\mathrm{e}_i.$$

The convergence proof in De Leeuw and Heiser (1980) implies that the current object space can be normalized to satisfy $\|\mathbf{X}_s\|^2 = 1$.

An update for $\mathbf{Y}_m$ is obtained as follows. First, from a starting point $\mathbf{Y}^o_m$, compute an unrestricted $N \times q_m$ matrix of scores $\tilde{\mathbf{Q}}_m$ (where the dimensionality $q_m \leq k_m$ may be chosen to vary over m if it is desirable to correct for possibly different numbers of categories):

$$\tilde{\mathbf{Q}}_m = N^{-1}\mathbf{B}(\mathbf{G}_m\mathbf{Y}^o_m)\mathbf{G}_m\mathbf{Y}^o_m,$$

where analogous to (6), the matrix $\mathbf{B}(\cdot) = \mathbf{B}^*(\cdot) - \mathbf{B}^o(\cdot)$, with elements

$$b^o_{ij}(\mathbf{G}_m\mathbf{Y}^o_m) = \frac{d_{ij}(\mathbf{X}_s)}{d_{ij}(\mathbf{G}_m\mathbf{Y}^o_m)} \qquad \text{if } m \in J_s,\ i{\neq}j, \text{ and } d_{ij}(\mathbf{G}_m\mathbf{Y}^o_m) \neq 0;$$

$$b^o_{ij}(\mathbf{G}_m\mathbf{Y}^o_m) = 0 \qquad \text{if } d_{ij}(\mathbf{G}_m\mathbf{Y}^o_m) = 0.$$

Note that the update does not involve the scalar factor a_{ms}, since we require $\mathbf{Y}_m$ to be normalized. The restrictions on $\mathbf{Y}_m$ can be separated: first an auxiliary $\tilde{\mathbf{Y}}_m$ that fits the categorical structure of $\mathbf{z}_m$ is found by simple projection: $\tilde{\mathbf{Y}}_m = (\mathbf{G}'_m\mathbf{G}_m)^{-1}\mathbf{G}'_m\tilde{\mathbf{Q}}_m$, and the normalized result, satisfying $\|\mathbf{G}_m\mathbf{Y}_m\|^2 = 1$, defines the update for $\mathbf{Y}_m$.

Once overall convergence has been obtained, each $\hat{\mathbf{Y}}_m$ can be rotated towards principal axes orientation. If $\mathbf{L\Lambda L}'$ denotes the eigenvalue decomposition of $\hat{\mathbf{Y}}'_m\mathbf{G}'_m\mathbf{G}_m\hat{\mathbf{Y}}_m$, the quantifications $\mathbf{Y}_m$ may be chosen as $\mathbf{Y}_m = \hat{\mathbf{Y}}_m\mathbf{L}$, so that consecutive dimensions in $\mathbf{G}_m\mathbf{Y}_m$ are orthogonal, and show differential saliences given by the sum of squares of $\mathbf{G}_m\mathbf{y}_{m(t)}$, where the subscript $m(t)$ denotes the t-th dimension in $\mathbf{Y}_m$.

Remark: In more conventional multidimensional scaling tasks, the fact that the gradient is not defined when $d_{ij} = 0$ (a problem not encountered in the majorization approach), could be viewed as being of merely theoretical interest. In the present scaling problem, however, because of the nature of the categorical data, there are many distances $d_{ij}(\mathbf{G}_m\mathbf{Y}_m) = 0$, and it becomes crucial that the algorithm still works, and can be proven convergent for distances equal to zero (De Leeuw 1977; De Leeuw and Heiser 1980).

References

DE LEEUW, J. (1977), "Applications of Convex Analysis to Multidimensional Scaling,"in *Recent Developments in Statistics,* Eds., J.R. Barra et al, Amsterdam: North-Holland, 133-145.

DE LEEUW, J., & HEISER, W.J. (1980), "Multidimensional Scaling with Restrictions on the Configuration,".in *Multivariate Analysis, Vol. V*, Ed., P.R. Krishnaiah, Amsterdam: North-Holland, 501-522.

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

GOWER, J.C. (1966), "Some Distance Properties of Latent Roots and Vector Methods Used in Multivariate Analysis," *Biometrika, 53*, 325-338.

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

HEISER, W.J., & DE LEEUW, J. (1977), *"How to Use SMACOF-I",* Research Report, Leiden, Leiden University: Department of Data Theory.

HUBERT, L.J. (1973), "Monotone Invariant Clustering Procedures," *Psychometrika, 38,* 47-62.

HUBERT, L.J. (1987), *Assignment Methods in Combinatorial Data Analysis,* New York: Marcel Dekker.

MAX, J. (1960), "Quantizing for Minimum Distortion," *Information Theory, 6,* 7-12.

MEULMAN, J.J, (1982), *Homogeneity Analysis of Incomplete Data,* Leiden: DSWO Press.

MEULMAN, J.J. (1986), *A Distance Approach to Nonlinear Multivariate Analysis,* Leiden: DSWO Press.

MEULMAN, J.J. (1992), "The Integration of Multidimensional Scaling and Multivariate Analysis with Optimal Transformations," *Psychometrika, 57,* 539-565.

MEULMAN, J.J., & VERBOON, P. (1993), "Points of View Analysis Revisited: Fitting Multidimensional Structures to Optimal Distance Components with Cluster Restrictions on the Variables," *Psychometrika, 58,* 7-35.

NISHISATO, S. (1980), *Analysis of Categorical Data: Dual Scaling and its Applications,* Toronto: University of Toronto Press.

NISHISATO, S. (1994), *Elements of Dual Scaling: An Introduction to Practical Data Analysis,* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

TUCKER, L.R (1951), *"A Method of Synthesis of Factor Analysis Studies",* Personnel Research Section Report No. 984, Washington, D.C.: Department of the Army.

TUCKER, L.R, & MESSICK, S. (1963), "An Individual Differences Model for Multidimensional Scaling," *Psychometrika, 28,* 333-367.

VAN RIJCKEVORSEL, J., BETTONVIL, B., & DE LEEUW, J. (1985), *"Recovery and Stability in Nonlinear Principal Components Analysis",* Research Report RR-85-21, Leiden, University of Leiden: Department of Data Theory.