RR-97-03

A DISTANCE-BASED VARIETY OF NONLINEAR MULTIVARIATE DATA ANALYSIS, INCLUDING WEIGHTS FOR OBJECTS AND VARIABLES

Jacques J.F. Commandeur
Patrick J.F. Groenen
Jacqueline J. Meulman

Department of Data Theory
Leiden University

A Distance-Based Variety of Nonlinear Multivariate Data Analysis, Including Weights for Objects and Variables

Jacques J.F. Commandeur, Patrick J.F. Groenen, and Jacqueline J. Meulman

December 22, 1997

Data Theory Group,

Department of Education,

Faculty of Social and Behavioral Sciences, Leiden University,

The Netherlands

This study is funded by The Netherlands Organization for Scientific Research (NWO) by grant nr. 575-67-053 for the 'PIONEER' project 'Subject Oriented Multivariate Analysis' to the third author.

Abstract

In the distance approach to nonlinear multivariate data analysis the focus is on the optimal representation of the relationships between the objects in the analysis. In this paper two methods are presented for including weights in distance-based nonlinear multivariate data analysis. In the first method, weights are assigned to the objects while the second method is concerned with differential weighting of groups of variables. When each analysis variable defines a group the latter method becomes a variable weighting method. For objects the weights are assumed to be given; for groups of variables they may be given, or estimated. These weighting schemes can also be combined and have several important applications. For example, they make it possible to perform efficient analyses of large data sets, to use the distance-based variety of nonlinear multivariate data analysis as an addition to loglinear analysis of multiway contingency tables, and to do stability studies of the solutions by applying the bootstrap on the objects or the variables in the analysis. These and other applications are discussed, and an efficient algorithm is proposed to minimize the corresponding loss function.

Keywords: object weights, group weights, optimal scaling, multivariate data analysis, multidimensional scaling, distance approach, preference data, bootstrap.

1. Introduction

In this paper two methods are presented to incorporate weights in the distance approach to nonlinear multivariate data analysis (MVA). These methods are concerned with the weighting of objects as well as of groups of variables. In the usual approach to nonlinear MVA (see, e.g., Gifi, 1990) the focus is on the optimization of the representation of the *variables* in the analysis. A disadvantage of this method is that the representation of the *objects* in the analysis can be suboptimal when viewed in terms of their mutual proximities. As has been discussed by De Leeuw and Meulman (1986) and Meulman (1986, 1992), one reason why this effect can arise is that standard MVA involves a projection onto low-dimensional space, thus yielding distances between the objects that are always smaller than their proximities. On the other hand, in the distance approach to nonlinear MVA the proximities between the objects in the data are directly approximated by distances in low-dimensional space. This approach is modeled by a loss function called STRESS in multidimensional scaling (Kruskal, 1964a, 1964b). Compared to standard nonlinear MVA, in nonlinear multivariate distance-based analysis (MVDA) the relations between the objects are by definition better represented in terms of their mutual proximities because this is the explicit objective of the analysis.

However, the latter approach also carries the computational burden of having to deal with $N(N-1)/2$ distances, where N is the number of objects. This burden can become prohibitive when the number of objects is very large. The use of *profile weights* may reduce the computational burden greatly, provided that the data matrix contains objects with (many) profiles that are equal to one another. The case of equal profiles allows for the analysis of the reduced data matrix only containing the $p < N$ distinct profiles by including the frequency counts of the profiles as weights in the analysis. This leaves only $p(p-1)/2$ distances to be taken into account.

When the data matrix consists of continuous variables, it is usually true that each row contains a different profile. In this case, one may consider to recode the variables into discrete categories. Hopefully, this would result in a (much) smaller number of distinct profiles which may then again be analysed (much) more efficiently by incorporating the profile counts as object weights in the analysis.

Previous work on profile weights in the context of a distance-based alternative for multiple correspondence analysis has been discussed by Groenen, Commandeur and Meulman (1996). They showed how to obtain an optimal representation of the objects when all the variables are categorical, a situation where optimal transformation of the variables is not required. In the present paper, the situation is discussed where the transformation level of the variables can be categorical, ordered categorical, or numerical, or where the variables are required to be transformed by spline functions, and where additionally these transformations of the variables have to be determined such that they are optimal in the context of MVDA with profile weights.

One important application of MVDA with profile weights is the analysis of multiway contingency tables where each cell represents a distinct profile, and each cell frequency represents a profile weight. In addition to standard loglinear analysis, MVDA of multiway contingency tables has the advantage that a graphical representation of the relations between all the cells of the table is obtained, while the order information of the categories is taken into account.

Apart from opening up the possibility for the analysis of large data sets and of multiway contingency tables, the use of object or profile weights also allows for a feasible implementation of the bootstrap method (Efron,1979) to study the stability of MVDA solutions (see, for an example, Groenen et al., 1996). Of course, irrespective of the size of the data set, the assignment of larger weights to some objects than to others can be used to give some objects a larger a priori influence on the solution.

In MVDA, we start from the general situation where the variables in the data matrix may be partitioned into two or more *groups*. Besides applying weights to objects or profiles in the data, in the present paper we additionally discuss two methods to assign weights to groups of variables in the data. When dealing with the situation where a group contains a single variable, such *group weights* become, in fact, *variable weights*, and defining each variable as a separate group therefore allows for differential weighting of the variables in the analysis.

In the first variable weighting method the weights are assumed to be given and fixed. In this case, weights can be used to give some (groups of) variables a more prominent role in the analysis than others. When dealing with one variable per group, it is very easy to implement a bootstrap for the variables to study the stability of the solution.

In the second weighting method the weights are estimated as part of the analysis. In this case, the value of the weight is a direct indication of the goodness-of-fit of the corresponding group of variables in the MVDA solution. Apart from yielding a very useful diagnostic, experience shows that the estimation of group weights is especially useful when the groups in the analysis contain different numbers of variables.

In section 2 we first show how object or profile weights can be incorporated in the MVDA loss function, and how to set up an efficient convergent algorithm for the minimization of this loss function. In section 3 variable weights are added to the loss function, and solutions are given for the corresponding minimization problems, both for fixed and estimated weights. Section 4 discusses a number of decompositions of the loss and fit in weighted MVDA. Then, in section 5, several illustrations of the weighting schemes proposed in the present paper are presented, as well as a stability study of MVDA with the bootstrap applied to the variables in the data matrix.

2. MVDA with Object or Profile Weights

Given is the $N{\times}m$ columnwise partitioned data matrix $\mathbf{Z} = [\,\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_M]$, containing M groups $\mathbf{Z}_J$ ($J = 1, \ldots, M$) of m_J variables each (thus, $m = \sum_J m_J$), with all variables defined on N objects. Each variable is assigned a transformation level which can be categorical, ordered categorical, or numerical, and variables may also be transformed by spline functions. In the usual loss function for MVDA the Euclidean distances $\mathrm{d}_{ik}(\mathbf{Q}_J)$ between objects i and k of group $\mathbf{Q}_J$ in the $N{\times}m$ *quantified* data matrix $\mathbf{Q} = [\,\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_M]$ are approximated by the Euclidean distance $\mathrm{d}_{ik}(\mathbf{X})$ between points representing objects i and k, and the STRESS loss function is written as

$$f(\mathbf{X},\mathbf{Q}) = \frac{1}{2}\sum_{J=1}^{M}\sum_{i=1}^{N}\sum_{k=1}^{N}(d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2 \qquad (1)$$

(see Meulman, 1992). This loss function has to be minimized over the $N \times s$ matrix of object scores $\mathbf{X}$ (with dimensionality s), and over the M quantifications matrices $\mathbf{Q}_J$ each of order $N \times m_J$.

When $\mathbf{Z}$ consists of only $p < N$ distinct profiles loss function (1) may also be evaluated as

$$f(\mathbf{X},\mathbf{Q}) = \frac{1}{2}\sum_{J=1}^{M}\sum_{i=1}^{p}\sum_{k=1}^{p} w_i w_k (d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2, \qquad (2)$$

where w_i denotes the profile weight for object i, which is the frequency count of the corresponding row. In the latter situation, matrices $\mathbf{Z}$, $\mathbf{Q}$, and $\mathbf{X}$ are assumed to contain only p rows. Since distances are unaffected by translations, we require without loss of generality that $\mathbf{X}'\mathbf{w} = \mathbf{0}$, and that $\mathbf{Q}'\mathbf{w} = \mathbf{0}$, where $\mathbf{w}$ denotes the $p \times 1$ vector containing the weights w_i. Moreover, to avoid the perfect but trivial solution $d_{ik}(\mathbf{Q}_J) = d_{ik}(\mathbf{X}) = 0$ for all i, k, and J, we impose the length restriction $\mathbf{q}_j'\mathbf{W}\mathbf{q}_j = \mathbf{w}'\mathbf{1}$ for all variables $j = 1, \ldots, m$, with $\mathbf{W}$ the diagonal matrix containing the p profile weights w_i on its diagonal, $\mathbf{1}$ the vector of ones, and $\mathbf{q}_j$ column j of the quantified data matrix $\mathbf{Q}$. When the weights w_i contain frequency counts of the profiles, we have that $\mathbf{w}'\mathbf{1} = N$. However, all results to be discussed below equally apply to the situation where each weight contains any nonnegative number; in the latter case we still assume the weights to add up to N, because they can always be normalized, without loss of generality, such that $\mathbf{w}'\mathbf{1} = N$. Expanding (2) gives

$$\begin{aligned} f(\mathbf{X},\mathbf{Q}) &= \frac{1}{2}\sum_{J=1}^{M}\sum_{i=1}^{p}\sum_{k=1}^{p} w_i w_k (d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2 \\ &= \textstyle\sum_J \sum_{i<k} w_i w_k d_{ik}^2(\mathbf{Q}_J) + M \sum_{i<k} w_i w_k d_{ik}^2(\mathbf{X}) - 2 \sum_J \sum_{i<k} w_i w_k d_{ik}(\mathbf{X}) d_{ik}(\mathbf{Q}_J) \\ &= \textstyle\sum_J \eta^2(\mathbf{Q}_J) + M\eta^2(\mathbf{X}) - 2 \sum_J \rho(\mathbf{X};\mathbf{Q}_J). \qquad (3) \end{aligned}$$

The minimization of (3) can not be solved analytically, but majorization theory (De Leeuw & Heiser, 1980; De Leeuw, 1988) can be applied to obtain a convergent algorithm for the minimization of f($\mathbf{X}$,$\mathbf{Q}$). This algorithm proceeds in two major steps, carried out iteratively, where the first step improves upon $\mathbf{X}$, keeping the quantifications matrices fixed, and the second step finds optimal transformations of the variables, keeping $\mathbf{X}$ fixed. Since (2) is bounded below by zero, this iterative procedure guarantees monotone convergence, and can be continued until solutions do not improve the value of the loss function beyond some small predetermined threshold value. Groenen et al. (1996) discuss the first step of this algorithm, and provide an efficient solution for the minimization of (3) over unrestricted matrix $\mathbf{X}$.

For fixed $\mathbf{X}$, it is shown in the Appendix how updates for the quantification matrices $\mathbf{Q}_J$ can be obtained efficiently by minimizing

$$g_J(*;\mathbf{Q}_J,\mathbf{R}_J) = d_J + N \operatorname{tr} (\mathbf{Q}_J - \mathbf{Q}_J^{\dagger})'\mathbf{W}(\mathbf{Q}_J - \mathbf{Q}_J^{\dagger}), \tag{4}$$

where d_J is a term independent of $\mathbf{Q}_J$, $\mathbf{R}_J$ denotes the previous optimal quantification matrix satisfying the constraints, and

$$\mathbf{Q}_J^{\dagger} = N^{-1}\mathbf{W}^{-1}\mathbf{B}(\mathbf{R}_J;\mathbf{X})\mathbf{R}_J. \tag{5}$$

We refer to the Appendix for a definition of matrix $\mathbf{B}(\mathbf{R}_J;\mathbf{X})$ in (5). To find optimal transformations for the variables in group $\mathbf{Q}_J$, that is, to minimize (4), we have to perform a weighted regression of each column in $\mathbf{Z}_J$ on the corresponding column in $\mathbf{Q}_J^{\dagger}$; the type of regression depends on the transformation level of the corresponding variable. For example, the regression for a numerical variable can be skipped, since linear transformations are fixed by the centering and length requirements imposed on the variables. For ordered categorical variables, (4) is a set of weighted monotone regression problems, while for categorical variables averages have to be calculated. Once optimal quantifications have been obtained, the variables of group J must all be normalized on $\mathbf{q}_j'\mathbf{W}\mathbf{q}_j = N$ (they are automatically centered on $\mathbf{q}_j'\mathbf{w} = \mathbf{0}$).

3. MVDA with Group Weights

If group weights a_J ($J = 1, \ldots, M$) are added to (2), and these group weights are assumed to be collected in the $M\times1$ vector $\mathbf{a}$, then the following loss function is obtained

$$f(\mathbf{X},\mathbf{Q},\mathbf{a}) = \sum_J \sum_{i<k} w_i w_k (a_J d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2. \tag{6}$$

A very similar problem has been discussed in Meulman (1993) in the context of points of view analysis. A group weight in (6) automatically becomes a variable weight when the corresponding group contains only one variable. In (6), the group weights a_J may either be fixed or free. To avoid the perfect but trivial solution $\{\mathbf{X} = \mathbf{0}, \mathbf{a} = \mathbf{0}\}$, when the group or variable weights are free we require, without loss of generality, the object scores $\mathbf{X}$ to satisfy tr $\mathbf{X}'\mathbf{W}\mathbf{X} = sN$. Whether the group weights are given or free, the update formula for $\mathbf{X}$ should be adapted to the following effect. The update for $\mathbf{X}$ becomes

$$\mathbf{X}^+ = \frac{1}{MN}\mathbf{W}^{-1}\sum_J a_J \mathbf{B}(\mathbf{X}^0;\mathbf{Q}_J)\mathbf{X}^0 \tag{7}$$

with matrix $\mathbf{B}(\mathbf{X}^0;\mathbf{Q}_J)$ as defined in the Appendix. Matrix $\mathbf{X}^0$ denotes the previous update for $\mathbf{X}$ satisfying $\mathbf{X}^{0\prime}\mathbf{w} = \mathbf{0}$ (see Groenen et al., 1996), and also satisfying tr $\mathbf{X}^{0\prime}\mathbf{W}\mathbf{X}^0 = sN$ if the group weights are free. In the latter free case, update (7) must be normalized on tr $\mathbf{X}^{+\prime}\mathbf{W}\mathbf{X}^+ = sN$. Due to the normalization restrictions on the quantification matrices $\mathbf{Q}_J$, no adaptations are required in formulas (4) and (5) for the minimization of (6) over the matrices $\mathbf{Q}_J$ when group weights are involved in the analysis.

If the group weights are free, a third step is added to the main MVDA algorithm in which updates for the group weights are calculated using

$$\hat{a}_J = \frac{\rho(\mathbf{X};\mathbf{Q}_J)}{m_J N^2} \quad \text{for } J = 1, \ldots, M. \tag{8}$$

These updates are found by setting the first derivative of (6) with respect to a_J equal to zero, and then solving for a_J. Because it is true that $\eta^2(\mathbf{X}) = N$ tr $\mathbf{X}'\mathbf{W}\mathbf{X} = sN^2$, and since also $\eta^2(\mathbf{Q}_J) = N$ tr $\mathbf{Q}_J'\mathbf{W}\mathbf{Q}_J = m_J N^2$, we have that

$$\frac{\rho(\mathbf{X};\mathbf{Q}_J)}{\eta(\mathbf{X})\eta(\mathbf{Q}_J)} = \frac{\rho(\mathbf{X};\mathbf{Q}_J)}{(sm_J)^{1/2}N^2} = (m_J/s)^{1/2}\,\hat{a}_J \equiv \tilde{a}_J\,. \tag{9}$$

This shows that the optimal group weights (8) are, in fact, equal to Tucker's congruence coefficient between the $p(p-1)/2$ elements $(w_i w_k)^{1/2} d_{ik}(\mathbf{X})$ and the $p(p-1)/2$ elements

$(w_iw_k)^{1/2}\mathrm{d}_{ik}(\mathbf{Q}_J)$, up to the constant factor $(m_J/s)^{1/2}$. Moreover, substitution of (8) in (6) shows that

$$\frac{1}{M}\sum_J \tilde{a}_J^2 + \frac{\mathrm{f}(\mathbf{X},\mathbf{Q},*)}{sMN^2} = 1, \tag{10}$$

meaning that the fit of each separate group of variables in the MVDA solution is equal to the squared Tucker's congruence coefficient, while the total fit equals the mean squared Tucker's congruence coefficient, averaged over groups.

As will be illustrated in section 5, the impact of estimated group weights on the solution of MVDA can be negligible when each group contains the same number of variables. However, the more different the number of variables per group, the larger the influence of estimated group weights on the solution becomes. The reason for this effect is that the sum of squares of each group of normalized variables is a direct function of the number of variables in the group: $\eta^2(\mathbf{Q}_J) = m_JN^2$ for $J = 1, \ldots, M$. Thus, group weights in MVDA will "correct" for these differences in sums of squares of the groups, and assign smaller weights to groups containing lots of variables, and larger weights to groups consisting of just a few variables.

Finally, we note that the normalization of update (7) for the object scores on tr $\mathbf{X}^{+\prime}\mathbf{W}\mathbf{X}^+ = sN$ is not required if the group weights a_J are fixed.

4. Diagnostics in MVDA

Because $\eta^2(\mathbf{Q}_J) = N$ tr $\mathbf{Q}_J'\mathbf{W}\mathbf{Q}_J = m_JN^2$, whatever the values in the quantification matrices $\mathbf{Q}_J$, we also have that $\sum_J a_J^2\eta^2(\mathbf{Q}_J) = N^2\sum_J m_J a_J^2$. Thus, for fixed group weights the minimum of $\mathrm{f}(\mathbf{X},\mathbf{Q})$ is the same as the minimum of $\mathrm{f}(\mathbf{X},\mathbf{Q})/\sum_J a_J^2\eta^2(\mathbf{Q}_J)$, because $\sum_J a_J^2\eta^2(\mathbf{Q}_J)$ is a constant. Whether the group weights are fixed or estimated, (6) can be written as $\sum_J a_J^2\eta^2(\mathbf{Q}_J) + M\eta^2(\mathbf{X}) - 2\sum_J a_J\rho(\mathbf{X};\mathbf{Q}_J)$. Because $M\eta^2(\mathbf{X}^+) \leq \sum_J a_J\rho(\mathbf{X}^+;\mathbf{Q}_J)$ (see De Leeuw, 1988), during the iterative process the ratio $\mathrm{f}(\mathbf{X},\mathbf{Q})/\sum_J a_J^2\eta^2(\mathbf{Q}_J)$ always satisfies $0 \leq \mathrm{f}(\mathbf{X},\mathbf{Q})/\sum_J a_J^2\eta^2(\mathbf{Q}_J) \leq 1$. Moreover, at convergence this ratio becomes equal to one minus the squared Tucker's congruence coefficient between the $p(p-1)/2$ elements $(w_iw_k)^{1/2}\mathrm{d}_{ik}(\mathbf{X})$ and the $p(p-1)/2$ elements $M^{-1}(w_iw_k)^{1/2}\sum_J a_J\mathrm{d}_{ik}(\mathbf{Q}_J)$. If the group weights are estimated, we have that $0 \leq f(\mathbf{X},\mathbf{Q},*)/sMN^2 \leq 1$, the latter ratio being equal to one minus the mean squared Tucker's congruence coefficient even during the iterative process (see (10)). These normalizations of the loss function are convenient, because they result in values that are independent of the number of objects, and of the values of the profile and the (fixed) group weights.

In the previous section, we already noted how estimated group weights are useful diagnostics for assessing the fit of the corresponding groups of variables in the MVDA solution. In the next two subsections a number of further interesting diagnostics are discussed, which are all based on decompositions of the total loss in $f(\mathbf{X},\mathbf{Q},\mathbf{a})$.

Primary Diagnostics

The total loss in $f(\mathbf{X},\mathbf{Q},\mathbf{a})$ indicates the overall goodness-of-fit of the solution. The total loss can be decomposed in such a way that the relative contributions of each separate group, each separate object, and also of each separate object in each separate group to the total loss can be evaluated. Specifically, the loss for object i in group J equals

$$f_{iJ} = \frac{MN}{2w_i} \sum_{k \neq i} w_i w_k (a_J d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2. \tag{11}$$

Averaging (11) over groups, the mean loss for object i becomes equal to

$$f_{i+} = \frac{N}{2w_i} \sum_J \sum_{k \neq i} w_i w_k (a_J d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2. \tag{12}$$

On the other hand, summing (11) over objects, and dividing by the total number of objects N, the mean loss in group J is obtained, which equals

$$f_{+J} = \frac{1}{N} \sum_i w_i \frac{MN}{2w_i} \sum_{k \neq i} w_i w_k (a_J d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2$$

$$= M \sum_{i<k} w_i w_k (a_J d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2. \tag{13}$$

If group weights are estimated using (8), (13) can also be written as $MN^2(s - m_J \hat{a}_J^2)$. Averaging (12) over objects, or averaging (13) over groups, we finally obtain

$$f_{++} = \sum_J \sum_{i<k} w_i w_k (a_J d_{ik}(\mathbf{Q}_J) - d_{ik}(\mathbf{X}))^2,$$

which is the total loss (6).

Secondary Diagnostics

If the group weights are fixed, at the point of convergence of the algorithm it is true that $M\eta^2(\mathbf{X}) = \sum_J a_J \rho(\mathbf{X};\mathbf{Q}_J)$; therefore, at convergence (6) can be written as

$$f(\mathbf{X},\mathbf{Q}) = N \sum_J a_J^2 \text{tr } \mathbf{Q}_J'\mathbf{W}\mathbf{Q}_J - NM \text{ tr } \mathbf{X}'\mathbf{W}\mathbf{X}, \tag{14}$$

from which it follows that

$$NM \text{ tr } \mathbf{X}'\mathbf{W}\mathbf{X} + f(\mathbf{X},\mathbf{Q}) = N\sum_J a_J^2 \text{tr } \mathbf{Q}_J'\mathbf{W}\mathbf{Q}_J = N^2\sum_J m_J a_J^2 , \tag{15}$$

and thus that

$$\frac{M \text{ tr } \mathbf{X}'\mathbf{W}\,\mathbf{X}}{N\sum_J m_J a_J^2} + \frac{f(\mathbf{X},\mathbf{Q})}{N^2\sum_J m_J a_J^2} = 1. \tag{16}$$

When the group weights are estimated, by substituting (8) into (6) it is not very difficult to show that (16) also applies. The left term in (16) represents the total fit of the solution, and can be further decomposed to indicate the overall contribution of each separate dimension to the fit of the solution, as well as the contribution of each separate variable to each separate dimension. The overall contribution of each separate dimension to the fit of the solution is equal to the diagonal elements of M diag$(\mathbf{X}'\mathbf{W}\mathbf{X})/N\sum_J m_J a_J^2$. At convergence of the algorithm the object scores are rotated to principal axes position, meaning that the contributions of the dimensions are equal to the eigenvalues of M diag$(\mathbf{X}'\mathbf{W}\mathbf{X})/N\sum_J m_J a_J^2$.

Let $\mathbf{G}_j$ of order $(p{\times}c_j)$ denote the indicator matrix of a categorical variable j, with c_j the number of categories of the variable, and let matrix $\mathbf{Y}_j = (\mathbf{G}_j'\mathbf{W}\mathbf{G}_j)^{-1}\mathbf{G}_j'\mathbf{W}\mathbf{X}$ of order $(c_j{\times}s)$ denote the weighted centroids of the object points corresponding to identical categories of the corresponding variable. Then the total fit can be decomposed into the following two independent components

$$N \text{ tr } \mathbf{X}'\mathbf{W}\mathbf{X} = N \text{ tr } \mathbf{Y}_j'\mathbf{G}_j'\mathbf{W}\mathbf{G}_j\mathbf{Y}_j + N \text{ tr } (\mathbf{G}_j\mathbf{Y}_j - \mathbf{X})'\mathbf{W}(\mathbf{G}_j\mathbf{Y}_j - \mathbf{X}), \tag{17}$$

which can also be expressed in terms of squared distances as

$$\sum_{i<j}^{n} w_i w_j \mathrm{d}^2_{ik}(\mathbf{X}) = \sum_{a<b}^{c_j} (\mathbf{g}'_{aj}\mathbf{W}\mathbf{g}_{aj})(\mathbf{g}'_{bj}\mathbf{W}\mathbf{g}_{bj})\mathrm{d}^2_{ab}(\mathbf{Y}_j) + N \sum_{k=1}^{c_j} \sum_{l=1}^{c_{kj}} w_l \mathrm{d}^2(\mathbf{x}_{lk}, \mathbf{y}_{jk}) \qquad (18)$$

where c_{kj} is the frequency of category k of variable j, $\mathbf{x}_{lk}$ is a profile point corresponding to category k of variable j, $\mathbf{y}_{jk}$ denotes the (weighted) centroid of category k of variable j, and $\mathbf{g}_{aj}$ is column a of indicator matrix $\mathbf{G}_j$. In words, (18) states that the weighted sum of squared distances of all profile points from the origin is equal to the weighted sum of squared distances of the centroids of the categories from the origin plus the weighted sum of squared distances of the profile points from the corresponding category centroids. It follows from (17) that the diagonal of (diag $\mathbf{X'WX})^{-1}$(diag $\mathbf{Y}_j'\mathbf{G}_j'\mathbf{W}\mathbf{G}_j\mathbf{Y}_j$) yields dispersion measures allowing one to assess the contribution of variable j to each dimension of the MVDA solution. Thus, if variable j discriminates the objects perfectly on all s dimensions, the discrimination measures add up to s; of course, the usual situation will be that they vary between zero and one on each dimension.

A special application of this idea is obtained if we construct cross-classified variables from the original variables (yielding a variable with as many categories as the product of the numbers of categories of the individual variables involved), and proceed as in (17). Such cross-classified variables make it possible to investigate the contributions of interaction variables to the fit of the solution (see the next section for an example).

5. Examples

The first example using object weights concerns the smoking attitude data described by Bull (1994). Two population surveys were conducted to evaluate the effect of the implementation of a bylaw in March 1988 regulating smoking in all workplaces in Toronto, Canada. The first survey with 1487 respondents was conducted in January-February 1988, just before the bylaw came into effect, and the second survey with 1368 respondents in November-December 1988, about 8 months later. One of the aims of the study was to investigate whether and how implementation of the bylaw in Toronto would affect the attitude of workers about smoking in the workplace. The data were obtained using telephone interviews. Each respondent in the study was assigned a so-called sampling weight according to the respondent's probability of selection from the household and according to the frequency of the respondent's age-sex category relative to the 1986 census distribution for metropolitan Toronto. This procedure was used by the researchers to ensure that the two independent surveys had the same age-sex distribution.

We used the variables described in Table 1 for an analysis with MVDA. The fifth and last variable in the table originally contained twelve categories, which we reduced to five using the discretization method described in Van Rijckevorsel, Bettonvil and De Leeuw (1985).

INSERT TABLE 1 ABOUT HERE.

The five variables in Table 1 consist of only 447 different profiles, and a weighted MVDA was performed on these 447 profiles in two dimensions; sampling weights were summed over identical profiles, and used as profile weights in the analysis. Except for the variable time, whose dichotomous character makes the type of transformation irrelevant, all variables were subjected to a monotone transformation. Differences in computing times between the weighted profile frequency matrix (n = 447) and the weighted full matrix (n = 2855) solutions are enormous. In this example, the computation of the solution based on the weighted profile frequency matrix is about 144 times faster than the one based on the weighted full matrix.

INSERT FIGURE 1 ABOUT HERE.

The plot of profile scores is given in Figure 1, labeled by the variables smoking attitude, education, smoking status, and knowledge. The bold characters in the plots are weighted centroids computed for the categories of the four interaction variables obtained by combining time with the other four variables. The most striking feature of the plots in Figure 1 is formed by the very distinct separation between the respondents interviewed before the implementation of the bylaw (lower cluster of profile points in the plots in Figure 1), and the respondents interviewed after the bylaw came into effect (upper cluster in the plots), showing that the second dimension is completely determined by the variable time. All the other variables differentiate between respondents on the first dimension. As the plots show, the first dimension differentiates the respondents on the left who never smoked, are well aware of the harmful effects of smoking, and are opposed to smoking in the workplace, from those respondents on the right who are current smokers, are not aware of harmful effects of smoking, and find that smoking in the workplace should be unrestricted. The variable education seems to have a nonlinear relation with the other variables. The stress for the separate variables is 0.429 for attitude, 0.361 for time, 0.442 for education, 0.362 for smoking status, and 0.396 for knowledge, with an average stress of 0.398. The overall conclusion of the analysis is that the implementation of the bylaw has had no effect whatsoever on the attitude of the respondents concerning smoking at work, and that this attitude is strongly related with their smoking status, and with what they know about the environmental effects of smoking.

In the second example we analyze the data in Table 2 from Sewell and Shah (1968), also discussed in Fienberg (1985), and consisting of five variables collected in a study of a randomly selected cohort of 10318 Wisconsin high school seniors. The five variables are: socioeconomic status (SES) with categories high, upper middle, lower middle, and low, intelligence (IQ) with the same four categories as variable SES, sex (with categories male and female), parental encouragement (categories low and high), and college plans (categories yes and no). The profile frequency matrix of these variables consists of only $4\times4\times2\times2\times2 = 128$ profiles, which may thus be efficiently analyzed using the frequencies of the profiles as profile weights. To the variables SES and IQ we applied spline transformations of degree 2 with 1 interior knot.

INSERT TABLE 2 ABOUT HERE.

Fienberg (1985) investigated the fit of a number of logit models to the college data using college plans as the response variable and all four remaining variables as explanatory. With logit models, the inherent order in the categories of the variables IQ and SES is not taken into account. To be able to compare our results with these models, we did a two-dimensional MVDA analysis of the college data storing variable college plans in one group, and the remaining four variables in a second group. The second reason that we thus distributed the five variables unevenly over two groups is that it allows for a demonstration of the differences that arise in such a situation when the estimation of group weights is added to the analysis. We first performed an analysis without estimating group weights. This yielded a solution with a stress of 0.200 for the group containing variable college plans, and a stress of 0.295 for the other group of variables, with an average stress of 0.2475. The plot of profile scores for this solution is shown in Figure 2.

INSERT FIGURE 2 ABOUT HERE.

We then repeated the analysis but additionally estimated group weights. The latter analysis yielded a solution with an average stress of .169, which is significantly better than the stress of .248 obtained in the analysis without group weights. The fit as calculated with (9) for the group containing variable college plans is .932, and that for the second group is .891. Thus, the overall fit of the solution equals $\frac{1}{M}\sum_J \tilde{a}_J^2 = 1/2\ (.932^2 + .891^2) = .831 = 1 - .169$ (see (10)). The result of the analysis is shown in Figure 3. Except that a better fit is obtained by including group weights, we also obtain a different solution. Compared to Figure 2, the distinction between the two categories of the separate variable college plans is much more pronounced.

INSERT FIGURE 3 ABOUT HERE.

Figure 4 contains the same profile scores as Figure 3, but now labeled with the categories of the variables in the second group: IQ, sex, parental encouragement, and SES. As can be clearly seen in the upper right plot in Figure 4, the second dimension of the solution completely separates the male high school seniors (upper part of the plot) from the female seniors (lower part of the plot). The remaining three variables IQ, parental encouragement, and SES primarily distinguish between the high school seniors on the first dimension of the solution, showing that college plans on the part of the seniors are associated with parental encouragement, as well as with the seniors being more intelligent and having a higher socioeconomic status, while having no plans to go to college is associated with not getting parental encouragement, as well as with lower intelligence and socioeconomical status. This is true for both male and female high school seniors.

INSERT FIGURE 4 ABOUT HERE.

Further decomposing the total fit of 1 – .169 = .831 in proportion of fitted scatter of the categories of each variable (see formula (17)), and computing $(\text{diag } \mathbf{X}'\mathbf{W}\mathbf{X})^{-1}(\text{diag } \mathbf{Y}_j'\mathbf{G}_j'\mathbf{W}\mathbf{G}_j\mathbf{Y}_j)$ the fit values in Table 3 are obtained. Summed over dimensions the dispersion measures of the variables are 0.821 for sex, 0.242 for intelligence, 0.946 for college plans, 0.478 for parental encouragement, and 0.279 for SES. Fitting his logit models to the college data, Fienberg (1985) found a model containing the first order interactions college plans and parental encouragement, and college plans and SES to provide a good fit to the data. To compare our results with those obtained in Fienberg, we constructed all first order interaction variables, and used (17) to compute the contributions of these interaction variables to the fit of the solution. We then computed departure from no first order interaction by computing the squared distance between the sum of the main effects vectors corresponding to the interaction, and the interaction vector calculated with (17). For example, the departure from no interaction for the first order interaction 'plans × SES' was computed as $((.946+.273) - .966)^2 + ((.000+.006) - .008)^2 = .064$. As Table 3 shows, the largest departures from no first order interaction are obtained for the same interactions found by Fienberg: college plans and parental encouragement, and college plans and SES.

The third example concerns preference data obtained from the 135 members of the Dutch parliament in 1990, and were previously discussed by Hillebrand and Meulman (1992). Each of the 135 parliament members was asked to assign a score between 0 and 100 to each of the six political parties that were represented in the same Dutch parliament. This score reflects the amount of sympathy that a member feels for a certain political party. Thus, a score between 0 and 49 reflects antipathy against a party (lower numbers reflecting more antipathy than higher numbers), and a score between 51 and 100 reflects sympathy for a party; score of fifty denotes a neutral attitude.

These data were analyzed with MVDA using the preference scores of the 135 parliament members as variables, and the six political parties as objects. To all 135 variables a quadratic spline transformation with one interior knot was applied. The stress of the MVDA solution is .2485. Including variable weights in the analysis yielded a stress of .2457, and almost exactly the same solution. Therefore, we only discuss the MVDA solution of the example without variable weights. The results of the analysis are shown in Figure 5. The small white squares represent the six parties. The first dimension discriminates between the denominational (SC and CDA) and the secular parties (GL, D66, and PvdA), with the VVD located somewhere in between. On the second dimension the parties are ordered according to importance in the Dutch parliament of 1990: at that time 48 members belonged to the PvdA, 46 to the CDA, 19 to the VVD, 11 to D66, 6 to Green Left, and 5 to the Small Christian parties.

Figure 5 also contains the 135 parliament members represented as vectors. These vectors were obtained by projecting the optimally transformed sympathy scores of each member into the object space using multiple regression. The vectors point in the direction of high sympathy and away from high antipathy. Figure 5 shows that parliament members usually sympathize most with their own party, as expected. For example, the bulk of the 46 vectors corresponding to the members of the CDA (labeled '4' in Figure 5) point in the direction of this party, and most of the 48 PvdA member vectors (labeled '2' in the figure) point towards the PvdA party point. However, there is one notable exception: The 19 vectors corresponding to the members of the VVD (labeled '5') point in quite a number of different directions. As the sympathy scores show, all but one of these members have the highest sympathy for their own party, but they all vary a lot on their remaining sympathy scores. Moreover, the mean stress of the parties in the MVDA solution are .19 for GL, .22 for the PvdA, .19 for D66, .25 for the CDA, .38 for the VVD, and .22 for the SC. This shows that, on the average, the 19 VVD members are the least well represented in the solution.

INSERT FIGURE 5 ABOUT HERE.

6. Stability Study

To study the stability of the MVDA solution for the 1990 Dutch parliament data, we applied the bootstrap (Efron, 1979) to the 135 variables in the analysis. For an introduction to the bootstrap, we refer to Efron and Tibshirani (1993). The bootstrap assesses stability by repeatedly sampling with replacement from the original data, analyzing these samples, and comparing the solutions. Since MVDA is primarily concerned with the relations between the objects in the analysis, in the present case the bootstrap is used to investigate the stability of the objects. For each object the bootstrap generates as many points as the number of samples. This cloud of points is then an empirical description of the distribution of the object. If the position of an object is stable, the cloud of points will show little dispersion. Here, we will study the stability of the object points under repeated sampling of the *variables* in the analysis. Bootstrapping the variables is particularly easy to implement in MVDA, since it is only the variable weights that differ between the bootstrap samples.

We applied the bootstrap to the 1990 Dutch parliament data. Each time, 135 variables were drawn randomly with replacement. Then, using the variable frequencies in the bootstrap sample as fixed variable weights, a MVDA solution was computed using the same options as for the analysis of the original Dutch parliament data (see section 5). This was repeated a 1000 times, yielding a 1000 configurations containing the six Dutch political parties.

Configurations obtained with MVDA are unique up to a translation, rotation, and reflection, since distances are invariant under these transformations. The scale of the configuration is also not very important since we are only concerned with relative distances. Moreover, when comparing two configurations, it is the ratio of distances that determines the congruence, not the actual scale. Therefore, to be able to compare the 1000 configurations obtained in the bootstrap study, these indeterminacies were removed by optimally translating, rotating, and dilating all 1000 bootstrap solutions to the target configuration obtained for the original Dutch parliament data. Thus, for each object 1000 bootstrap points were obtained. Instead of displaying all 1000 bootstrap points for each of the six political parties under study, we computed an elliptical region that contains 95% of the bootstrap sample points (Meulman & Heiser, 1983).

INSERT FIGURE 6 ABOUT HERE.

Figure 6 shows the 95% confidence regions for every object based on the 1000 bootstrap samples from the 135 variables of the 1990 Dutch parliament data, together with the object scores of the original solution, and the centroids of the 95% confidence regions. Clearly, the six objects of the original sample fall well within the 95% regions, showing that the present MVDA solution is very stable. Another proof for the stability of the solution is that the confidence regions are quite small, and that the regions are well separated, except for the small overlap between the regions for GL and D66. Finally, the object points of the original solution are located very near the centroids of the confidence regions, which is an indication that the estimated object scores are not biased (see Efron and Tibshirani, 1993).

7. Discussion

We have proposed a method for the analysis of multivariate data based on distances between profiles, where the data may contain variables that are categorical, ordered categorical, or numerical, as well as variables to be transformed by spline functions. This method can be viewed as a special case of multidimensional scaling. By using frequency counts of profiles as object weights, MVDA is extended to deal efficiently with large data sets, if the number of profiles is much smaller than the number of objects; an example of such data is a multiway contingency table with large frequencies in the cells.

MVDA is also extended with a weighting scheme which allows for the weighting of groups of variables. When group weights are estimated, they both yield a useful diagnostic for the evaluation of the fit of the variable groups in the solution, and correct for differential contributions due to different numbers of variables in the groups. Fixed group weights, on the other hand, make it possible to control the amount of influence of individual variables on the MVDA solution.

Moreover, both weighting schemes can be used to investigate the stability of the object scores of MVDA solutions with the bootstrap. With object weights this can be done by sampling the objects, with fixed group weights by sampling the variables. For the 1990 Dutch parliament data, the bootstrap study showed that the MVDA solution is very stable under random sampling of the 135 variables. However, further study is needed to corroborate this result.

8. Appendix

In this appendix is discussed why and how new efficient updates for the quantification matrices $\mathbf{Q}_J$ can be obtained by minimizing (4) (see section 2). Thus, we assume $\mathbf{X}$ in (3), that is, in

$$\mathrm{f}(\mathbf{X},\mathbf{Q}) = \frac{1}{2} \sum_{J=1}^{M} \sum_{i=1}^{p} \sum_{k=1}^{p} w_i w_k (\mathrm{d}_{ik}(\mathbf{Q}_J) - \mathrm{d}_{ik}(\mathbf{X}))^2,$$

to be fixed, and show how to solve the second step of the main algorithm: How to find optimally transformed variables $\mathbf{Q}$, that are also centered and normalized. First, we note that (3) may be written in matrix form as

$$\mathrm{f}(*,\mathbf{Q}) = M\eta^2(\mathbf{X}) + \textstyle\sum_J \mathrm{tr}\ \mathbf{Q}_J'\mathbf{V}\mathbf{Q}_J - 2\sum_J \mathrm{tr}\ \mathbf{Q}_J'\mathbf{B}(\mathbf{Q}_J;\mathbf{X})\mathbf{Q}_J, \tag{19}$$

where matrix $\mathbf{B}(\mathbf{Q}_J;\mathbf{X})$ has off-diagonal elements $b_{ikJ} = -w_i w_k \mathrm{d}_{ik}(\mathbf{X})/\mathrm{d}_{ik}(\mathbf{Q}_J)$ if $\mathrm{d}_{ik}(\mathbf{Q}_J) \neq 0$, $b_{ikJ} = 0$ if $\mathrm{d}_{ik}(\mathbf{Q}_J) = 0$, and diagonal elements $b_{iiJ} = -\sum_{k \neq i} b_{ikJ}$. Matrix $\mathbf{V}$ has off-diagonal elements $v_{ik} = -w_i w_k$, and diagonal elements $v_{ii} = -\sum_{k \neq i} v_{ik}$, and the rank of $\mathbf{V}$ is $(p - 1)$.

Applying majorization theory, a restricted update for $\mathbf{Q}$ is obtained by using a combination of majorization and projection. Specifically, only considering one group of variables $\mathbf{Q}_J$, and letting $\mathbf{R}_J$ denote the previous optimal quantification matrix satisfying the constraints, we have to minimize the majorization function

$$g_J(*,\mathbf{Q}_J,\mathbf{R}_J) = \eta^2(\mathbf{X}) + \text{tr } \mathbf{Q}_J'\mathbf{V}\mathbf{Q}_J - 2 \text{ tr } \mathbf{Q}_J'\mathbf{B}(\mathbf{R}_J;\mathbf{X})\mathbf{R}_J$$

$$= c_J + \text{tr } (\mathbf{Q}_J - \mathbf{Q}_J^{\dagger})'\mathbf{V}(\mathbf{Q}_J - \mathbf{Q}_J^{\dagger}), \qquad (20)$$

where $c_J = \eta^2(\mathbf{X}) - \text{tr } \mathbf{Q}_J^{\dagger\prime}\mathbf{V}\mathbf{Q}_J^{\dagger}$ is a term independent of $\mathbf{Q}_J$, and

$$\mathbf{Q}_J^{\dagger} = \mathbf{V}^{-}\mathbf{B}(\mathbf{R}_J;\mathbf{X})\mathbf{R}_J, \qquad (21)$$

with $\mathbf{V}^{-}$ a generalized inverse of $\mathbf{V}$. The convergence proofs of De Leeuw and Heiser (1980) show that the metric projection in (20) of $\mathbf{Q}_J^{\dagger}$ in the space satisfying the constraints will never increase the value of loss function (2). However, when p is large the computation of the inverse of $\mathbf{V}$ needed in (21) is quite a task. Moreover, projection problem (20) has to be solved in the complicated metric $\mathbf{V}$.

Fortunately, all this can be avoided by using the following much more efficient procedure. As mentioned in Groenen et al. (1996), matrix $\mathbf{V}$ has the simple structure

$$\mathbf{V} = N\mathbf{W} - \mathbf{w}\mathbf{w}', \qquad (22)$$

where $\mathbf{w}$ and $\mathbf{W}$ are defined as in section 2. Since we require that $\mathbf{Q}_J$ has a weighted column mean of zero (i.e., that $\mathbf{Q}_J'\mathbf{w} = \mathbf{0}$, see section 2), (22) allows us to write (20) as

$$g_J(*;\mathbf{Q}_J,\mathbf{R}_J) = \eta^2(\mathbf{X}) + N \text{ tr } \mathbf{Q}_J'\mathbf{W}\mathbf{Q}_J - 2 \text{ tr } \mathbf{Q}_J'\mathbf{B}(\mathbf{R}_J;\mathbf{X})\mathbf{R}_J,$$

$$= d_J + N \text{ tr } (\mathbf{Q}_J - \mathbf{Q}_J^{\dagger})'\mathbf{W}(\mathbf{Q}_J - \mathbf{Q}_J^{\dagger}),$$

where $d_J = \eta^2(\mathbf{X}) - N \text{ tr } \mathbf{Q}_J^{\dagger\prime}\mathbf{W}\mathbf{Q}_J^{\dagger}$ is again a term independent of $\mathbf{Q}_J$, and

$$\mathbf{Q}_J^{\dagger} = N^{-1}\mathbf{W}^{-1}\mathbf{B}(\mathbf{R}_J;\mathbf{X})\mathbf{R}_J.$$

Thus, the computation of the inverse of $\mathbf{V}$ is reduced to the computation of the inverse of diagonal matrix $\mathbf{W}$, and the metric projection problem is reduced to the simple diagonal metric $\mathbf{W}$.

For the interested reader, a FORTRAN program is available on the internet (http://www.fsw.leidenuniv.nl/www/w3_data/pioneer/pioneer.htm) in which all the options discussed in the present paper have been implemented.

References

Bull, S. (1994). Analysis of attitudes toward workplace smoking restrictions. In N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest, & J. Greenhouse (Eds.), *Case studies in biometry* (pp. 249-271). New York: Wiley.

De Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification*, *5*, 163-180.

De Leeuw, J. & Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In P.R. Krishnaiah (Ed.), *Multivariate analysis, Vol. V* (pp. 501-522). Amsterdam: North-Holland.

De Leeuw, J. & Meulman, J.J. (1985). *Alternative approximations in principal components analysis.* Research Report RR-85-13. Leiden: Department of Data Theory.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, *7*, 172–184.

Efron, B. & Tibshirani, R.J. (1993). An introduction to the bootstrap. New York: Chapman & Hall.

Fienberg, S.E. (1985). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: M.I.T. Press.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester: Wiley.

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

Hillebrand, R. & Meulman, J.J. (1992). Afstand en nabijheid: verhoudingen in de Tweede Kamer. In J.J.A. Thomassen, M.P.C.M. van Schendelen, M.L. Zielonka-Goei (Eds.), *De geachte afgevaardigde... hoe kamerleden denken over het Nederlandse parlement.* Muiderberg: Coutinho.

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

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29,* 115-129.

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

Meulman, J.J. (1993). Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika, 58,* 7-35.

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

Meulman, J.J. & Heiser, W.J. (1983). *The display of bootstrap solutions in multidimensional scaling.* Unpublished manuscript.

Sewell, W.H. & Shah, V.P. (1968). Social class, parental encouragement, and educational aspirations. *American Journal of Sociology, 73,* 559-572.

TABLE 1

Variables of Smoking Attitude Data of Bull (1994)

1)	attitude toward smoking in the workplace: smoking should be 1 – prohibited 2 – restricted 3 – unrestricted	2)	time of survey relative to implementation of by-law 1 – pre 2 – post
3)	education 1 – elementary 2 – some high school 3 – high or trade school 4 – college or some university 5 – university degree	4)	smoking status 1 – never smoked 2 – quit > 12 months ago 3 – quit 6-12 months ago 4 – quit ≤ 6 months ago 5 – current smoker
5)	knowledge of health effects of environmental tobacco smoke (ranges from 1 to 5)		

TABLE 2

The College Data of Sewell and Shah (1968)

Sex	IQ	College Plans	Parental Encouragement	SES L (1)	LM (2)	UM (3)	H (4)
M (1)	L (1)	Yes (1)	Low (1)	4	2	8	4
			High (2)	13	27	47	39
		No (2)	Low (1)	349	232	166	48
			High (2)	64	84	91	57
	LM (2)	Yes (1)	Low (1)	9	7	6	5
			High (2)	33	64	74	123
		No (2)	Low (1)	207	201	120	47
			High (2)	72	95	110	90
	UM (3)	Yes (1)	Low (1)	12	12	17	9
			High (2)	38	93	148	224
		No (2)	Low (1)	126	115	92	41
			High (2)	54	92	100	65
	H (4)	Yes (1)	Low (1)	10	17	6	8
			High (2)	49	119	198	414
		No (2)	Low (1)	67	79	42	17
			High (2)	43	59	73	54
F (2)	L (1)	Yes (1)	Low (1)	5	11	7	6
			High (2)	9	29	36	36
		No (2)	Low (1)	454	285	163	50
			High (2)	44	61	72	58
	LM (2)	Yes (1)	Low (1)	5	19	13	5
			High (2)	14	47	75	110
		No (2)	Low (1)	312	236	193	70
			High (2)	47	88	90	76
	UM (3)	Yes (1)	Low (1)	8	12	12	12
			High (2)	20	62	91	230
		No (2)	Low (1)	216	164	174	48
			High (2)	35	85	100	81
	H (4)	Yes (1)	Low (1)	13	15	20	13
			High (2)	28	72	142	360
		No (2)	Low (1)	96	113	81	49
			High (2)	24	50	77	98

TABLE 3

Fit of Main Variables and First Order Interaction Variables in MVDA Solution of College Data, and Departure of First Order Interaction Variables from No First Order Interaction

	Fit by dimension		Total fit
	1	2	
MAIN VARIABLES			
sex	.011	.810	.821
IQ	.224	.019	.242
plans	.946	.000	.946
encour	.473	.005	.478
SES	.273	.006	.279
FIRST ORDER INTERACTION VARIABLES			
sex × IQ	.235	.855	1.090
sex × plans	.947	.818	1.764
sex × encour	.474	.814	1.288
sex × SES	.283	.833	1.117
IQ × plans	.959	.022	.981
IQ × encour	.553	.034	.587
IQ × SES	.394	.026	.420
plans × encour	.983	.009	.992
plans × SES	.966	.008	.975
encour × SES	.544	.021	.565
DEPARTURE FROM NO FIRST ORDER INTERACTION			
sex × IQ			.001
sex × plans			.000
sex × encour			.000
sex × SES			.000
IQ × plans			.044
IQ × encour			.021
IQ × SES			.011
plans × encour			.190
plans × SES			.064
encour × SES			.041

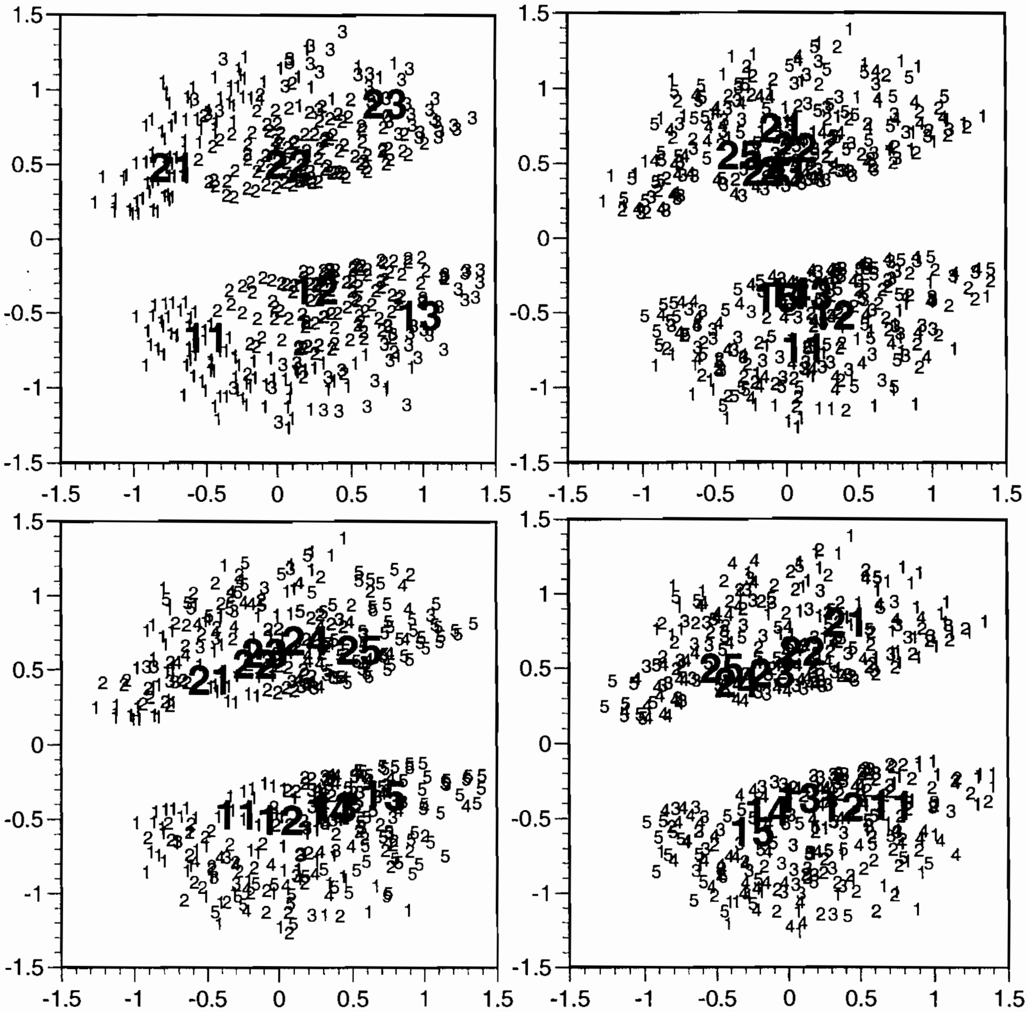

Figure 1. Plot of profile scores obtained with MVDA of smoking attitude data, labeled by attitude (upper left), education (upper right), smoking status (lower left), and knowledge of health effects of environmental tobacco smoke (lower right). The lower cluster of points in all plots consists of the respondents interviewed before the bylaw came into effect, the upper cluster of all respondents after implementation of the bylaw. The numbers in bold in the four plots are weighted centroids of the categories of the cross-classified variables time/attitude, time/education, time/smoking status, and time/knowledge, respectively.

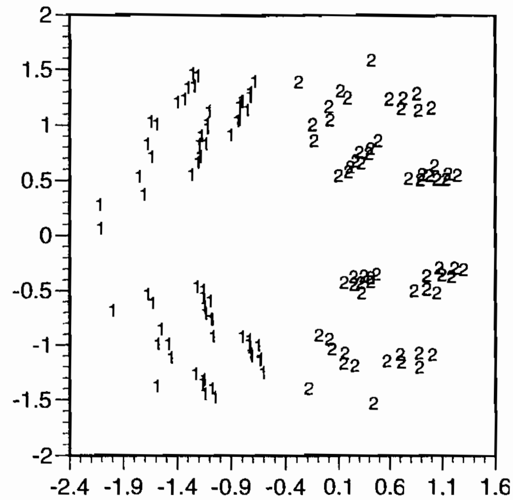

Figure 2. Plot of profile scores of college data obtained with MVDA without estimating group weights, labeled by college plans. The profile points on the left represent the high school seniors who plan to go to college, the points on the right consist of all seniors having no such plans.

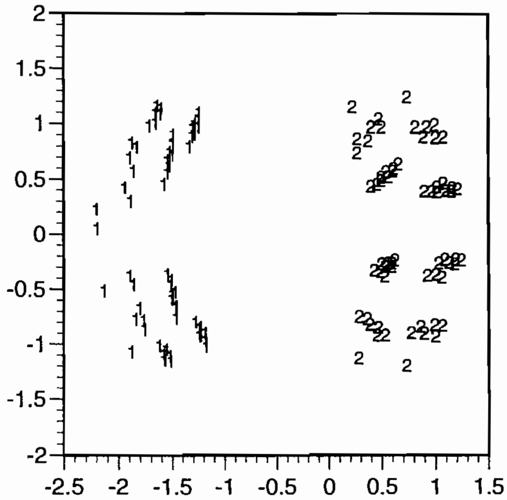

Figure 3. Plot of profile scores of college data obtained with MVDA with group weights, labeled by college plans. The profile points labeled with a '1' represent the high school seniors planning to go to college, those labeled with a '2' represent the seniors not planning to go to college.

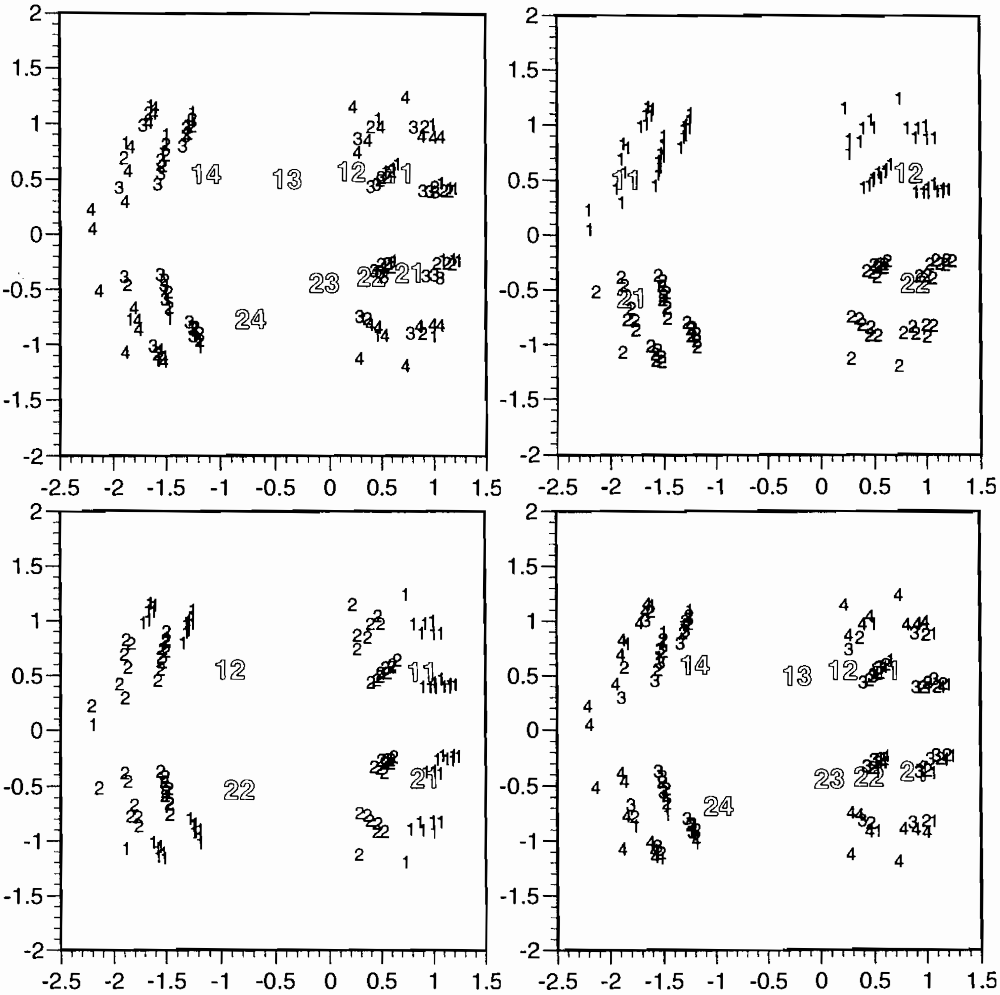

Figure 4. Plot of profile scores obtained with MVDA with free group weights of college data, labeled by IQ (upper left), sex (upper right), parental encouragement (lower left), and socio-economic status (lower right). The numbers in outline in the four plots are weighted centroids of the categories of the cross-classified variables sex/IQ, sex/college plans, sex/parental encouragement, and sex/socio-economic status, respectively.

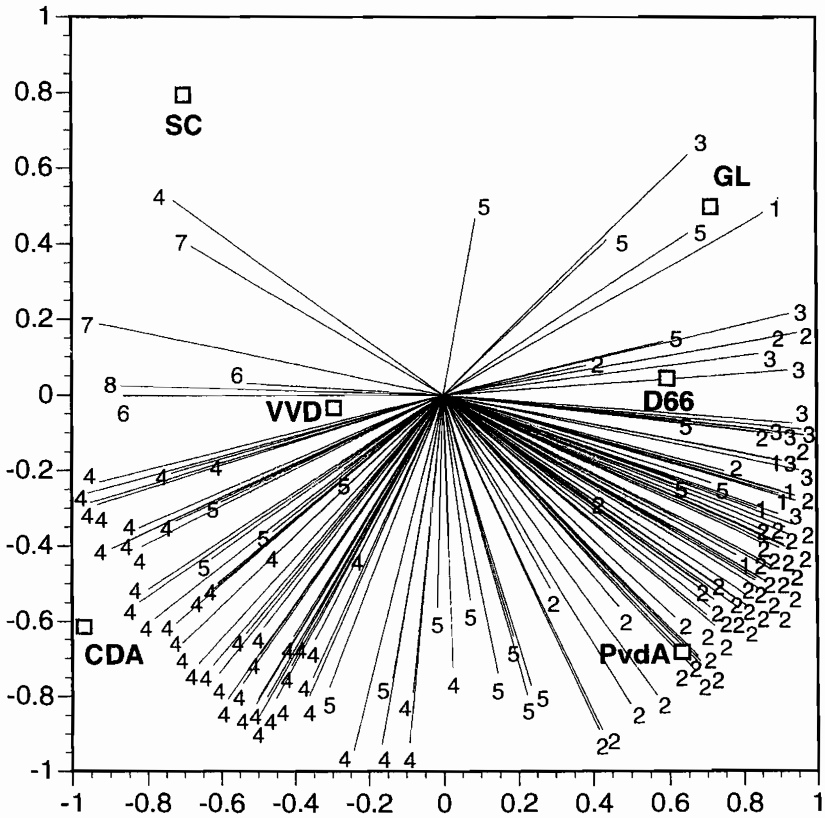

Figure 5. MVDA solution of 1990 Dutch parliament preference data: 1 = GL (Green Left), 2 = PvdA, 3 = D66, 4 = CDA, 5 = VVD, 6 = GPV, 7 = RPF, 8 = SGP, SC = Small Christian parties (GPV, RPF, and SGP).

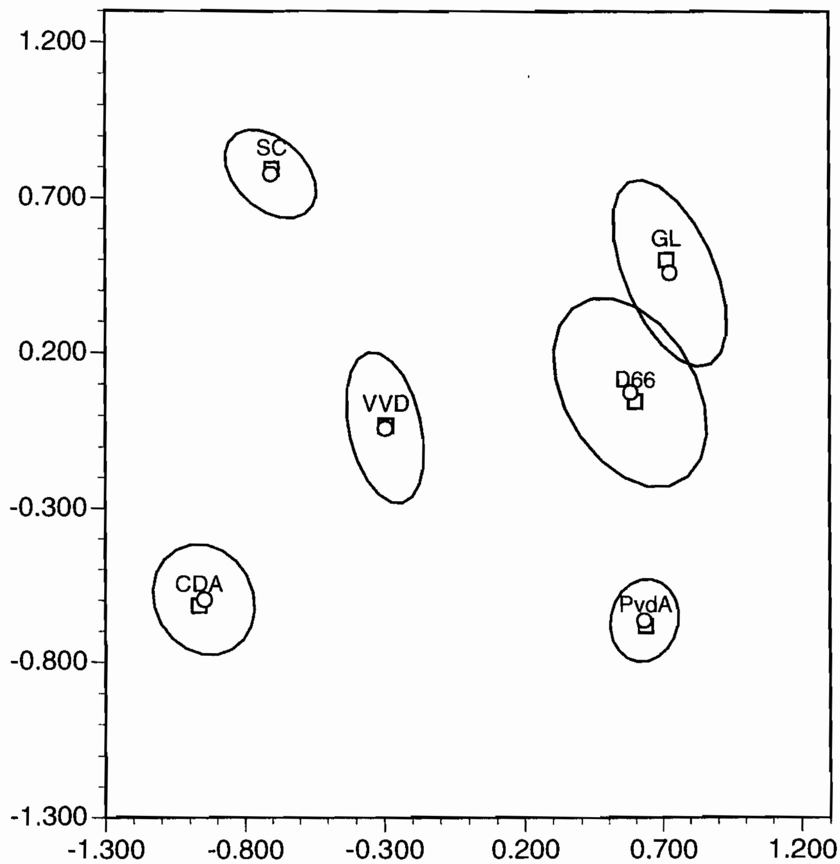

Figure 6. Plot of MVDA 95% confidence regions for 1000 bootstraps on the 135 Dutch parliament preference score variables; the small squares denote the original object scores, the small circles the centroids of the confidence regions.